Total Exposure Health
An Introduction

Edited by
Kirk A. Phillips, Dirk P. Yamamoto, and
LeeAnn Racz

ISBN: 978-0-367-20139-5 (hbk)
ISBN: 978-0-429-26328-6 (ebk)

First published 2020

Chapter 16
Exposure Health Informatics Ecosystem

Ramkiran Gouripeddi, Philip Lundrigan, Sneha Kasera, Scott Collingwood, Mollie Cummins, Julio C. Facelli, and Katherine Sward

The chapter DOI: 10.1201/9780429263286-18

The funder of the Open Access version of this chapter is University of Utah

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

16 Exposure Health Informatics Ecosystem

Ramkiran Gouripeddi
University of Utah

Philip Lundrigan
Brigham Young University

Sneha Kasera, Scott Collingwood, Mollie Cummins, Julio C. Facelli, and Katherine Sward
University of Utah

CONTENTS

16.1 DETERMINANTS OF HEALTH

While estimates of the exact contributions vary between studies, at least 50% of a person's health can be attributed to their environment, lifestyle, and behavior (Centers for Disease Control and Prevention 2019, Tarlov 1999, McGinnis, Williams-Russo, and Knickman 2002, Choi and Sonin 2019). A study in England attributed 40% of all disease burden to identifiable risk factors and almost 75% of these to

DOI: 10.1201/9780429263286-18

a combination of individuals' environmental, behavioral, and metabolic profiles (Newton et al. 2015). About one in four deaths worldwide, and a similar proportion of deaths among children under five, are due to modifiable environmental factors (Prüss-Üstün et al. 2016, Landrigan et al. 2018). These 12.6 million deaths are attributed to more than 100 different diseases (World Health Organization 2016a,b).

By itself, air pollution is linked to one in eight deaths globally (World Health Organization 2014, 2017). About two billion children live in areas exceeding the World Health Organization's annual limits for fine particles (United Nations International Children's Emergency Fund 2016). It is noted that 169,250 and 531,000 child deaths are attributable to ambient and household air pollution, respectively (World Health Organization 2017). The global social cost of air pollution is about $3 trillion per year (Erickson and Jennings 2017). Recent studies have provided strong evidence associating air quality with pediatric asthma (Pollock, Shi, and Gimbel 2017). At the same time, there is discussion about our environment being cleaner than ever before and the unprimed nature of our immune systems being responsible for disease manifestation (Richtel 2019). While it is reasonably well understood that continuous exposure to high levels of pollution is unhealthy, much less is known about health effects of low levels, intermittent exposure, and combinations of exposures. Studying the latter requires informatics infrastructures that can aggregate environmental and physiological data from multiple sources at high temporal and spatial resolutions. Research and development of such multi-scale and multi-model informatics infrastructure is just beginning; and in this chapter, we describe the requirement, design, and development activities undertaken to address these issues.

16.2 THE EXPOSOME AND ITS GENERATION

Comprehensive quantification of effects of the modern environment on health requires taking into account data from all contributing environmental exposures and how those exposures relate to health; this is termed the *exposome*, a complementary concept to the genome (Wild 2005, 2012). Measuring the exposome can span a lifetime of exposures starting from conception and includes endogenous processes within the body, biological responses of adaptation to environment, physiological manifestations of these responses, and socio-behavioral factors (Wild 2005, 2012). Generating exposomes at high resolution requires integration of data from wearable and stationary sensors, environmental monitors, personal activities, physiology, medication use and other clinical data, genomic and other biospecimen-derived, person-reported and computational models. Exposomic research is translational in nature, as the exposome includes direct biological pathway alterations, as well as mutagenic and epigenetic mechanisms of environmental influences on the phenome (Miller 2013, Miller and Jones 2014, Lioy and Weisel 2014, NIOSH 2018). The *phenome*, which is an individual's state of well-being and disease, is a result of the interaction between a person's *genome* and their exposome. See Figure 16.1 for a holistic understanding of disease, integration of the exposome, genome, and other factors.

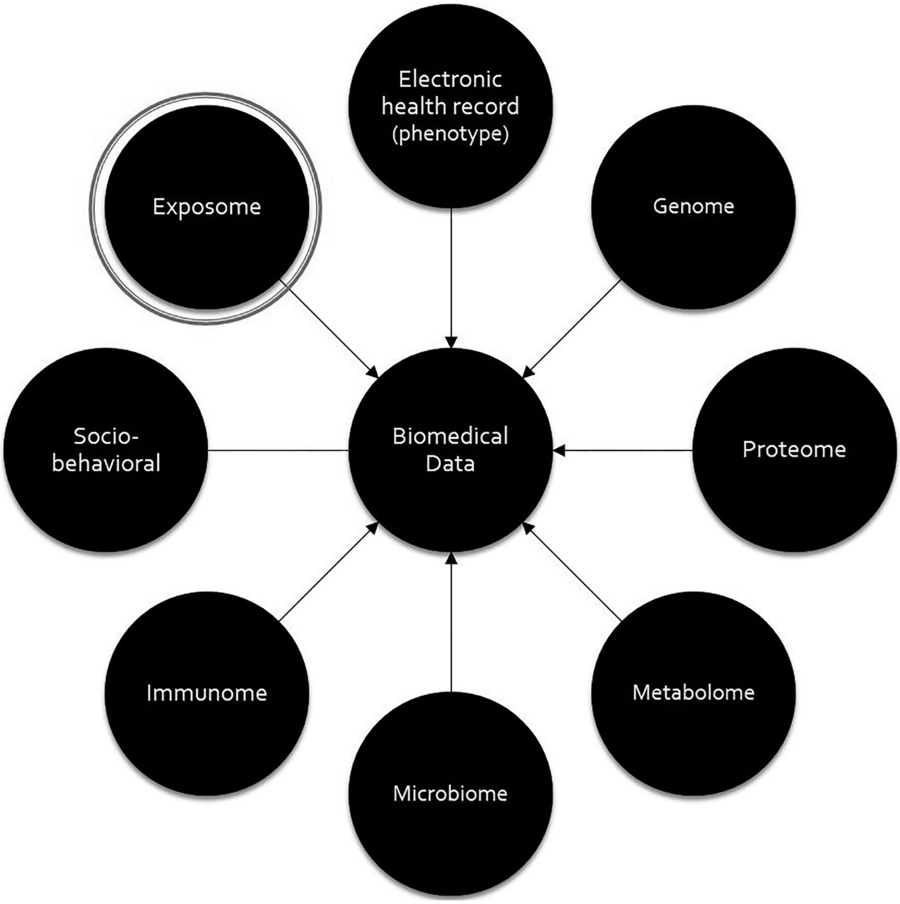

FIGURE 16.1 Holistic understanding of disease requires integration of the exposome with the genome with other biomedical data.

There is a need for understanding an individual's total exposure including simultaneous, cumulative, and latent exposure to multiple environmental species on health (Pollock, Shi, and Gimbel 2017). We refer to any physical (e.g., temperature, humidity), chemical (e.g., particulate matter (PM), ozone), or biological (e.g., pollen, mold) environmental or physiological (e.g., breath rate, forced expiration volume) entity measured by a sensor as a species. Processes to support this aggregation and integration must accommodate variable spatio-temporal resolutions and account for multiple study, experimental and analytical designs. Gaps in measured data may need to be filled with modeled data along with characterization of uncertainties.

The air quality exposome is important to our improved understanding of pediatric asthma and other respiratory conditions (Pollock, Shi, and Gimbel 2017), cardiovascular disease (Lee, Kim, and Lee 2014), cancers (Santibáñez-Andrade et al. 2017), pregnancy (Leiser et al. 2019), suicide (Bakian et al. 2015, Gładka, Rymaszewska, and Zatoński 2018), and its mechanistic role in damage to deoxyribonucleic acid (Bosco et al. 2018, Miri et al. 2019). It includes a combination of chemical (PM, ozone, and volatile organic compounds), biological (pollen, spores) and physical (temperature, humidity) environmental species. Studies involving the exposome can be observational, epidemiological, interventional, or mechanistic in nature (Röhrig et al. 2009).

16.3 THE PEDIATRIC RESEARCH USING INTEGRATED SENSOR MONITORING SYSTEMS PROGRAMS

The Pediatric Research using Integrated Sensor Monitoring Systems (PRISMS) program was launched in 2015 to develop a sensor-based, data-intensive infrastructure for measuring environmental, physiological, and behavioral factors for performing pediatric and adult epidemiological studies (https://www.nibib.nih.gov/research-funding/pediatric-research-using-integrated-sensor-monitoring-systems). PRISMS is administered by the National Institutes of Health (NIH) National Institute of Biomedical Imaging and Bioengineering (NIBIB). Figure 16.2 shows the various projects under the PRISMS program.

The University of Utah was funded through this program to identify informatics challenges and develop solutions to address them (http://prisms.bmi.utah.edu/). Recognizing that solving these challenges will require a wide range of perspectives, the Utah team is a diverse group of faculty, research staff, software developers, post-doctoral fellows, and graduate and undergraduate students from atmospheric science, bioengineering, biomedical informatics, chemical engineering, chemistry, clinical and translational science, computer science, electrical and computer engineering, industrial engineering, nursing, occupational health, pediatrics and pulmonary medicine.

16.4 EXPOSOMIC RESEARCH CHALLENGES AND INFORMATICS SOLUTIONS

Exposomic research requires simultaneous measurement of many types of environmental, physiological, and behavioral factors using sensors. These measurements can be obtained using sensor technologies that are often novel and in various stages of development, evolving to capture measurements of novel species, with improvements in their sensitivity, performance, and validity in measuring different species, in their form factor so that they can be used in personal and mobile settings, and price. In addition, these sensors use diverse device communication protocols and require additional hardware and software modifications for using research studies and for secure data acquisition and transmittal.

Environmental species have spatial and temporal variations and humans are mobile and spend time at home, commuting, at work or school, and in recreation.

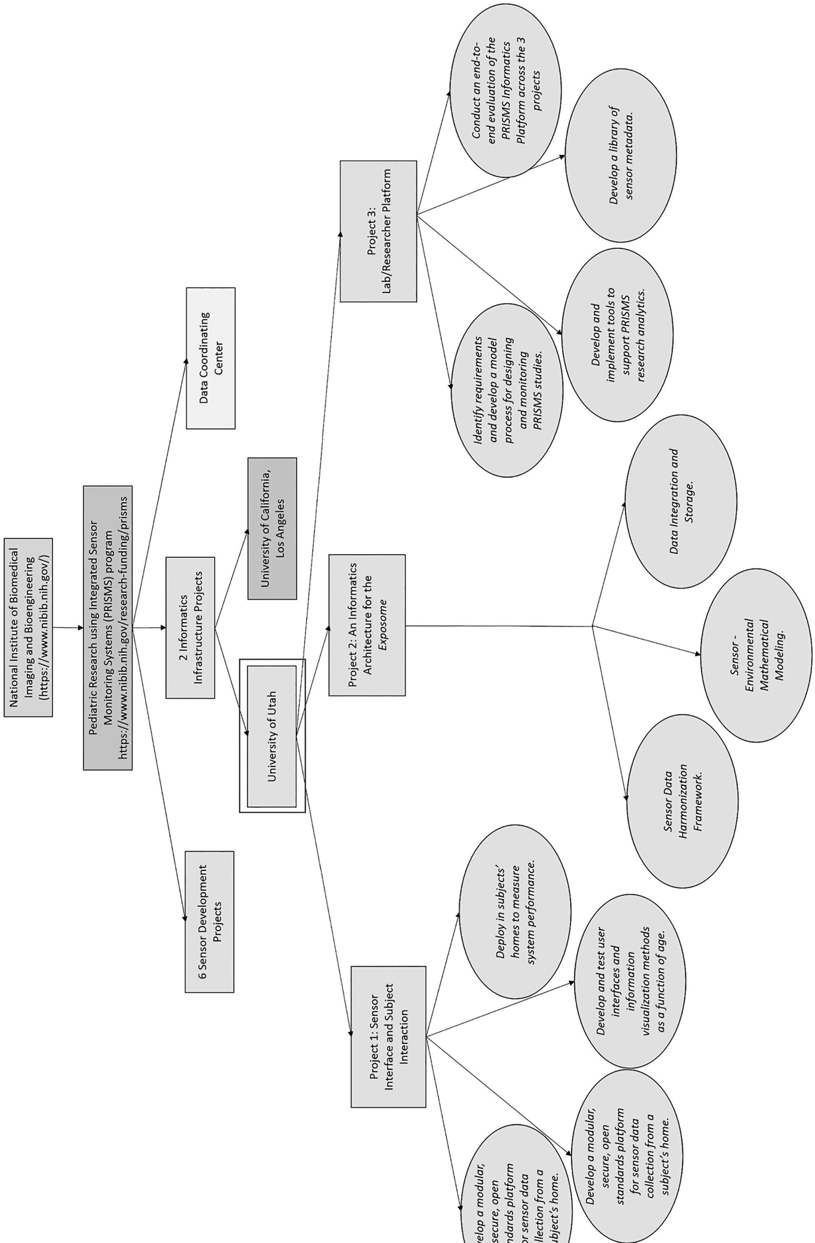

FIGURE 16.2 The PRISMS program and tasks being performed by the University of Utah.

Generation of comprehensive spatio-temporal records of exposures requires collection and integration of data from different types of sensors that might be available at different locations and times corresponding to the locations of the subject under consideration (Gouripeddi et al. 2017). For example, an air quality exposome may require the integration of data from indoor and mobile sensors, stationary regulatory monitors, citizen's networks, and finally supplementation with data from computational models to fill in the gaps when there is an absence of experimental data. All of these would require appropriate spatio-temporal dimensions and resolution with their absence often limiting the quality of studies and potentially leading to erroneous results (Gouripeddi et al. 2017).

Moreover, sensors used for measurements of exposures are not always collocated with the subject under consideration. The lack of proximity of the sensor to the subject leads to uncertainties when using their measurements as exact quantifications of exposures. In addition, sensors have varied capabilities, granularities and resolutions in measuring different environment species, which need to be harmonized prior to analysis.

Total exposure research studies need to be performed across health conditions, age ranges, and sensor types and utilize heterogeneous data at multiple levels of granularities in their semantics and temporalities. Different translational research archetypes require different data, data transformations, data integration workflows, and analytics to support observational and interventional study designs (Gouripeddi 2016).

Addressing these challenges in exposomic research requires an informatics architecture that embeds multiple features that are loosely coupled and interoperable (Sward et al. 2017, Martin Sanchez et al. 2014):

1. Sensor data acquisition: The evolving nature of sensors requires a sensor data acquisition paradigm that is agnostic to the sensor and the type of the species it is measuring. In addition, acquiring these sensor data should accommodate mobile and stationary devices that measure personal and ambient environments.
2. Selection of heterogeneous data sources: Prospective studies require use of sensors that are well-matched for the purposes of the study. Secondary analyses require descriptions about sources and methods including types of sensors used, to support appropriate analysis. In both cases, research teams require metadata about the sensors and the data sources.
3. Computational modeling for filling gaps: It may not be possible to measure every environmental variable at the desired temporal and spatial resolution, either due to availability and/or challenges with use of sensors, cost, privacy, number of sensors needed in large cohort studies, etc. Having computational models to help fill gaps can provide substitutes for or augment sensor-measured environmental factors, activities, and locations of individuals.
4. Uncertainty characterization of data: Understanding limitations and data quality of sensors, their measurements and, similarly, computational models

would enable their proper use within data pipelines, designing appropriate studies, and performing apt analysis. These limitations and uncertainties can be captured and shared as metadata.

5. Generation of a high-resolution spatio-temporal grid of exposures: Exposures are intrinsically tied to location and time. Different sources of exposure data, sensors, and computational model need to be combined to generate a high-resolution grid of personal exposure. These sources could have different granularities and resolution, and their integration would need to support these heterogeneity.
6. Data integration: To support the above requirements, integration of these heterogeneous exposomic data would need to be semantically consistent (Habre et al. 2016) and metadata driven. In addition, the diversity of different objects represented in translational exposomic research require them to be integrated on their spatial and temporal dimensions. Representing data as *events* permits temporal analysis and reasoning around a diverse array of environmental measurements, physiological responses, and conditions. An event-based infrastructure would support multi-scale and multi-omics integration.
7. Presentation and visualization: In order to make meaningful use of the data and processes in exposomic research, there is a need for acceptable and user-friendly interfaces for study participant and investigator interactions. These interfaces will provide feedback, allow participants to be provided instructions for interventions, and a means for participants to input additional requested data. Investigators will be able to manage study processes, assess ongoing data collections, and tailor interventions. There will likely be a need to have these presentation and visualization layer be person-centered and on mobile platforms.
8. Support a diverse set of translational research archetypes: The informatics infrastructure would need to support diverse study types, including observational, epidemiological, interventional, secondary analysis, and mechanistic study. In addition the infrastructure would need to enable reproducibility and transparency of study results with metadata to track data and process provenances.

16.5 EXPOSURE HEALTH INFORMATICS ECOSYSTEM

In order to meet the diverse requirements listed above, we are developing a scalable informatics infrastructure, Exposure Health Informatics Ecosystem (EHIE) (Sward and Facelli 2016, Sward et al. 2017, Gouripeddi et al. 2019b) following an ecosystemic approach. An ecosystem is a collection of loosely coupled software and hardware platforms that co-evolve, interact with one another and with human actors to serve a common business need (i.e., research, in this case), by maintaining symbiotic operational relationships with each other through exchange of data, metadata, knowledge, and process artifacts (Messerschmitt and Szyperski 2003, Jansen, Finkelstein, and Brinkkemper 2009, Lungu 2009, Popp and Meyer 2010, Jansen, Brinkkemper, and Cusumano 2013, Bala Iyer 2014). Adopting this approach enables researchers to sustain healthy large-scale infrastructures, by having a diversity in tasks performed by different actors within the ecosystem (Manikas and Hansen 2013). In addition, having this diversification in niche components simplifies the management and evolution of the ecosystem as a whole, as each component has its own development cycle managed by a team of experts (Dittrich 2014). Modifications from a specific component can be independently scaled as needed for a particular research use case. Similarly, operational use of the ecosystem paradigm for research studies would have support from several actors with appropriate expertise together providing greater value than on their own (Wnuk et al. 2014).

EHIE is derived from the federally funded National Institutes of Health's (NIH) National Institute of Biomedical Imaging and Bioengineering (NIBIB) Pediatric Research Using Integrated Sensor Monitoring Systems (PRISMS) program (Sward et al. 2016). EHIE addresses the above list of exposomic research challenges by providing informatics solutions at scale, incorporating the latest *Big Data* approaches. The infrastructure is a comprehensive, standards-based, open-source informatics platform that provides semantically consistent, metadata-driven, event-based management of exposomic data. Using an event-driven architecture allows the modeling and storage of all activities related to the study itself and its operations in their primitive form on a timeline as events that can be transformed to higher/analytical models based on use cases. Moreover, its implementation using advanced graph and document store technologies limits semantic dissonance and enables the use of novel Big Data approaches in a natural way. EHIE is aligned with the goals of modern environmental health research supporting meaningful integration of sensor and biomedical data (National Institute of Environmental Health Sciences 2012, Barksdale Boyle et al. 2015, National Institute of Environmental Health Sciences 2018). See Figure 16.3 for an overview of EHIE informatics ecosystem.

Conceptually, all the evolving software and hardware artifacts within EHIE can be grouped into the following components:

1. *Data acquisition pipeline*: Hardware and software tools, wireless networking, and protocols to support easy sensor system deployment and robust sensor data collection.
2. *Participant-facing tools*: Collect and annotate a variety of patient-reported and activity data, as well as inform and provide feedback to study participants on their current clinical and environmental exposure status.
3. *Researcher-facing platforms*: Tools and processes for researchers undertaking exposomic studies for a variety of experimental designs or for clinical care.
4. *Computational modeling platform*: Generate comprehensive spatio-temporal data in the absence of measurements and for recognition of activity signatures from sensor measurements.

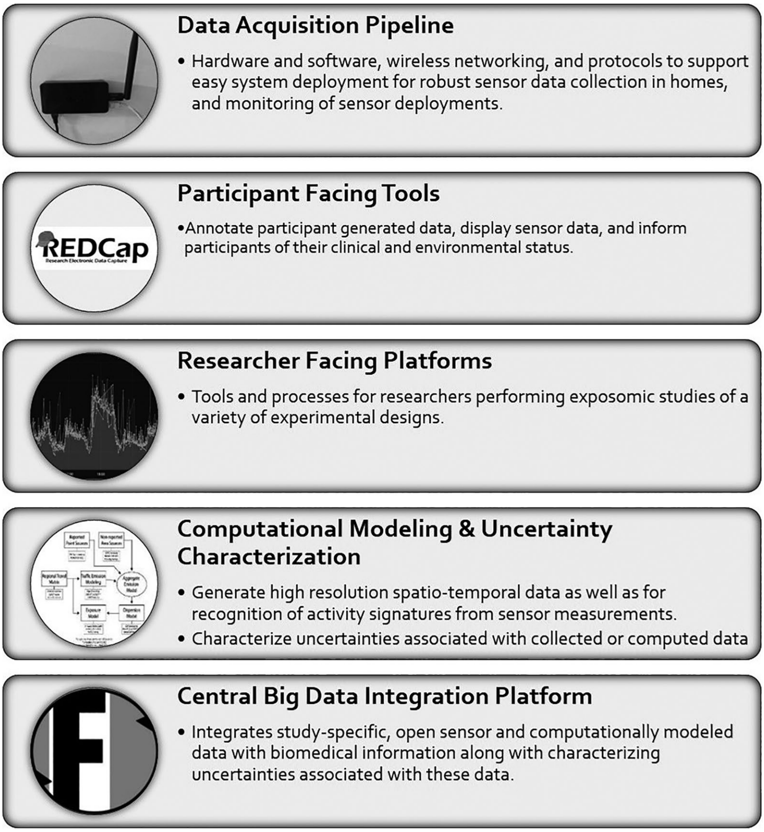

FIGURE 16.3 Exposure Health Informatics Ecosystem (EHIE) and its main components.

5. *Central Big Data federation/integration platform*: Standards-based, open-access infrastructure that integrates measured and computationally modeled data with biomedical information along with characterizing uncertainties associated with using these data.

In the following sections, we describe key features of each of these components.

16.5.1 Data Acquisition Pipeline

Current *Internet-of-Things (IoT)* solutions are not necessarily designed for health research. Systems are not designed for large study-based deployments wherein the cost and resources required for management of IoT sensors exceeds the cost of the sensors themselves. Research solutions are required to be compliant with pertinent privacy laws applicable at different jurisdictions (e.g., deployment site(s), study site, and/or study sponsor location) for data transmission (Luxton, Kayl, and

Mishkind 2012) and storage. While IoT sensors provide low-cost and smart solutions to measure study participants' environments, they usually use custom software and hardware, require regular maintenance, and have data integrity problems. We, therefore, needed to design an open-source platform that is customizable to different sensors, study designs, and participant requirements. Such a platform would need to have a short deployment time, provide high-quality data, and stream data in real time enabling control loops of feedback and interventions.

In order to meet these needs, we developed a multi-pronged approach for data acquisition. We developed EpiFi (Figure 16.4) (Lundrigan et al. 2018) to overcome these limitations and extend the use of off-the-shelf sensor technologies as IoT solutions for health research. In addition, we developed methods and processes for sensors that can directly transmit data to data acquisition servers, using protocols such as the Message Queuing Telemetry Transport (MQTT) (Hunkeler, Truong, and Stanford-Clark 2008) or HTTP/HTTPS. Our collaborators at Columbia University have adopted AethLabs sensors (www.aethlabs.com) to use these protocols to measure and transmit measurements for PM composition, black carbon, temperature and relative humidity, accelerometry and volatile organic compounds levels (Cox et al. 2019). In order to help investigators choose an appropriate approach, we developed a framework that considers the type of sensors and their transmission, study design, and participant involvement (Tiase et al. 2018). EpiFi provides flexibility in using existing participant home infrastructure and accommodates participant-in-the-loop study designs.

EpiFi brings IoT to health research by providing robustness to consumer applications needed by different study designs. It allows researchers, participants and their families, and clinicians to process data in real time. It simplifies the process of IoT deployment and management in hundreds of participant homes as might be needed in clinical studies. EpiFi consists of a small single-board computer (i.e., Raspberry Pi) gateway and open-source Home Assistant home automation platform (Home Assistant 2019), with custom code to address challenges of using sensors for research data acquisition. It has means to reliably transfer to a remote database using a home WiFi router

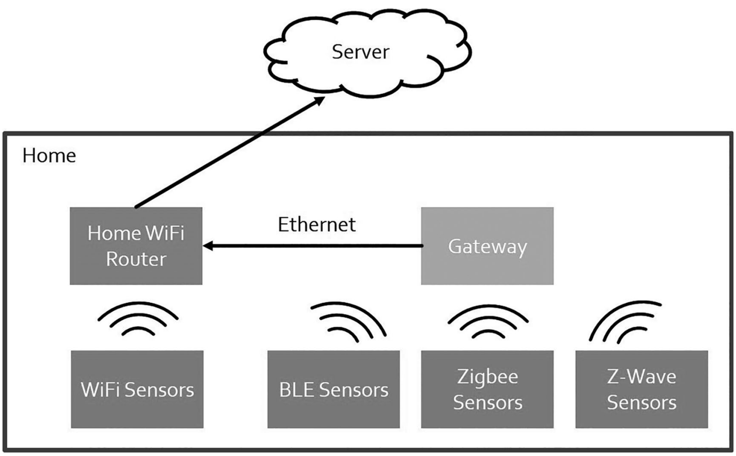

FIGURE 16.4 Overview of EpiFi.

and local storage that can act as a buffer when transmission to the remote database is not available or required. The system architecture of EpiFi is shown in Figure 16.5.

EpiFi (Lundrigan et al. 2018) supports multiple features that make it appropriate for use as an IoT solution in clinical studies:

1. *Device observability*: Allows a remote study manager to know if a WiFi device is functioning or not. It distinguishes between WiFi disruptions and other types of disruptions, so that appropriate troubleshooting can be performed.
2. *Secure WiFi bootstrapping*: Allows secure bootstrapping of WiFi connectivity of multiple devices by making the gateway a temporary access point. By overloading the use of source and destination addresses of an Ethernet frame, the Secure Transfer of Association Protocol (STRAP) (Lundrigan, Kasera, and Patwari 2018) allows a trusted device on the network to send data to unconnected WiFi devices (Figure 16.6). This protocol addresses the challenges of securely connecting new sensors within a home. It protects against eavesdroppers, modified messages, replay attacks, and rogue access point attacks. STRAP also reduces deployment time by needing the entry home WiFi credentials only on EpiFi and eliminating the need of entry by each individual sensor.

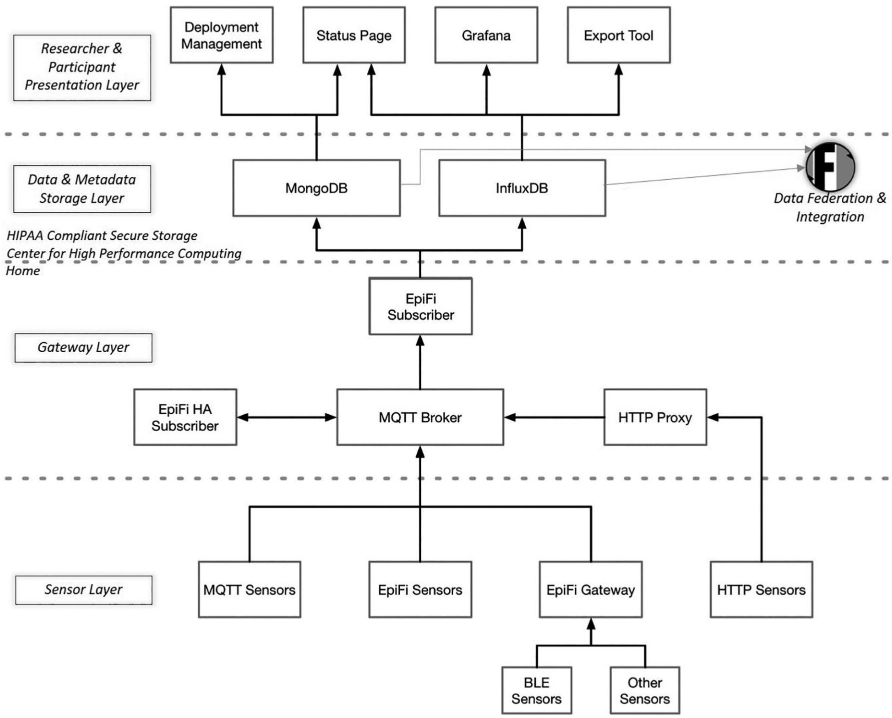

FIGURE 16.5 Architecture of EpiFi.

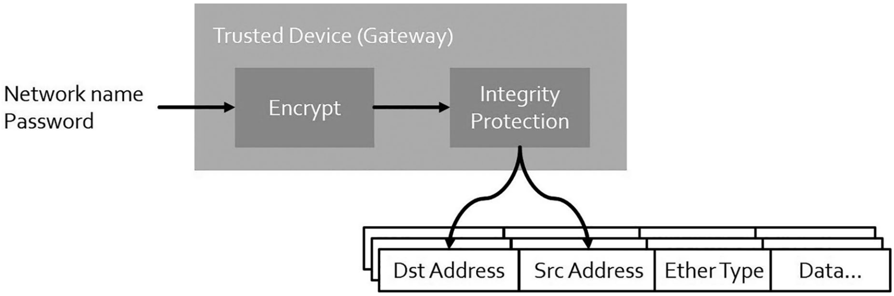

FIGURE 16.6 The Secure Transfer of Association Protocol (STRAP).

3. *Secure sensor reuse*: Tracking of sensors when their location changes and management of backlogged data on sensors. Sensors learn their locations based on network characteristics. A change in network characteristics indicates that the sensor is now in a new location, which then sets off processes to update deployment metadata. Also, each location has a key that is used to encrypt data, which prevents backlogged data from being read by a person at a different location.
4. *Study management tools*: Provides a presentation layer to support a diverse range of tools for study management (Figure 16.5). Integrated bi-directional communications with the gateway device help with remote management and troubleshooting potential sources of signal disruption and apply fixes. We currently use the following study management tools for the PRISMS pilot study (Collingwood et al. 2018, Gouripeddi et al. 2019c):
    a. *Deployment status page*: Provides the status of various deployments.
    b. *Export tool*: Export streaming data in various formats for ad hoc analysis.
    c. *Grafana (2019)*: Streaming data visualization, monitoring, and analytics.
5. *Support of multiple wireless protocols*: Supports among other cellular, Z-Wave (Yassein, Mardini, and Khalil 2016), ZigBee (Farahani 2011), LoRa (Lee and Ke 2018), and HaLow (802.11ah) (Adame et al. 2014).
6. *Data integrity*: Prevents data loss arising due to packet losses, gateway outages, home WiFi router outages, and internet outages by persisting data at every opportunity, deleting persisted data only after acknowledgment of receipt from remote storage, and sending multiple data packets when backlogged.

EpiFi is currently deployed for pilot studies at Utah for facilitating acquisition of data from:

1. WiFi sensors
    a. *Utah Modified Dylos (UMD) (Min et al. 2018, Vercellino et al. 2018, Collingwood et al. 2019)*: PM as PM2.5 (i.e., PM which is 2.5 µm and smaller) and PM10 (i.e., 10 micrometers and smaller), temperature, and humidity.
    b. *AirU (Kelly et al. 2017)*: PM.

2. Bluetooth Low Energy (BLE) sensors
   a. *Wearable air quality sensor from George Washington University (Li et al. 2019)*: Nitrogen dioxide (NO_2), ozone (O_3), ambient temperature, formaldehyde, other aldehydes, and relative humidity.
   b. *Wearable air quality sensor from Arizona State University (Wang and Tao 2017)*: O_3, volatile organic compounds (VOCs), ambient temperature, relative humidity, accelerometry, nitrogen oxides (NOx), formaldehyde (CH_2O), and PM.
   c. *Wearable device from the University of Maryland (Chatterjee et al. 2014, Kukkapalli et al. 2016)*: PM, temperature, transcutaneous partial carbon dioxide (CO_2) pressure, and respiratory rate.

EpiFi has been evaluated in different types of deployment designs, including:

1. *High-resolution air sensing (Min et al. 2018)*: EpiFi acquired data from eight UMDs and AirUs deployed indoors and outdoors, respectively, to create a profile of air quality within a home due to various activities. For example, one of our findings for a home under study showed that, while the furnace fan rapidly improves PM levels in the kitchen, there were short-term increases in PM in other rooms.
2. *Automation of interventions (Min et al. 2018)*: We demonstrated that EpiFi can be integrated into home automated control systems, such as a furnace fan, via an Ecobee thermostat which triggers the furnace fan to switch on when PM levels crossed a present threshold measured by UMDs. This smart control of the furnace fan led to a 70% reduction in power consumption when compared to periodically turning it on.
3. *Requisition of clinical status, feedback, and activity annotation (Collingwood et al. 2018)*: Using EpiFi, we were able to send text notifications to participants when specific thresholds of PM levels were crossed, to acquire participant clinical status, feedback, and log the activities they were performing.
4. *Acquisition heterogeneous sensor data from participant homes*: Including motion sensors, door sensors, tracking smartphones, participant locations, smart light bulbs, WiFi usage, temperature, humidity, energy meters, and any other commercial IoT device to get a sense of participant activities.

Data from EpiFi is stored remotely in a time-series database (Figure 16.5), i.e., InfluxDB (InfluxData 2019). Metadata about the deployments is authored through a graphical user interface into a deployment metadata repository (DMDR) that has been instantiated in a MongoDB database (MongoDB 2019). Together these two stores, along with a set of Software Services (SS), support the presentation layer which provides displays that can be used by participants, researchers, and additional administrative tools. A tracking page was developed to provide the real-time health status of each deployed sensor allowing the administrative team to detect, analyze, and troubleshoot issues in various deployments using established procedures and protocols (Figure 16.7).

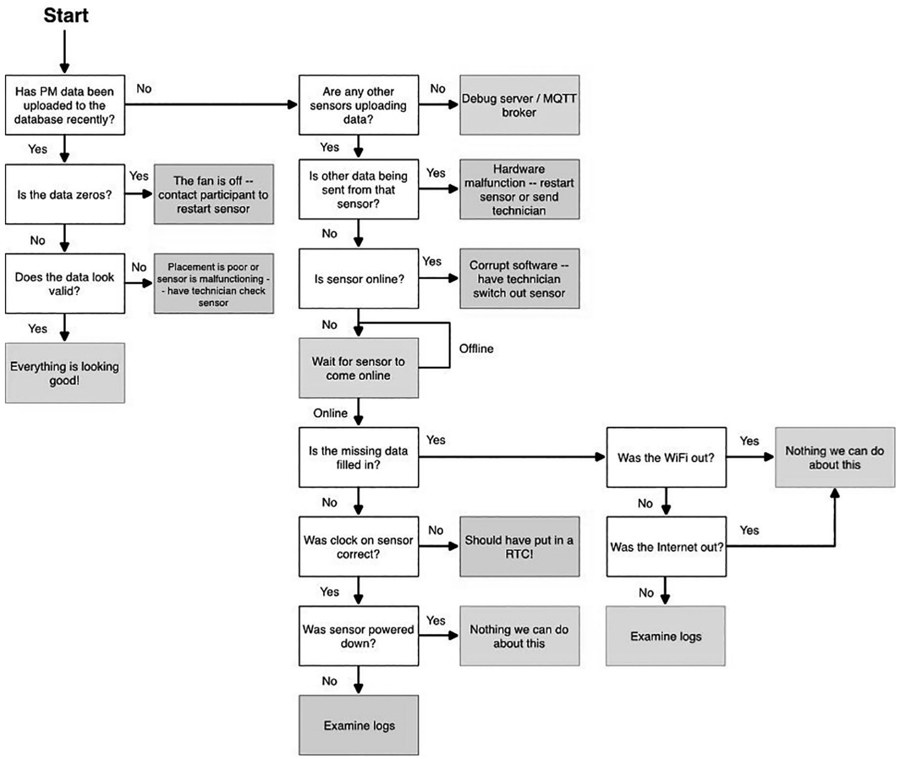

FIGURE 16.7 Example troubleshooting protocol.

In addition to supporting the presentation layer, data and metadata from the time-series database and the DMDR are consumed by the SS of the data federation and integration component for assimilation, generation of exposure records, and study analysis. Software details on EpiFi are available in Lundrigan (2019). We are currently evaluating blockchain approaches for systematically capturing the versioning of sensor deployment metadata as sensors go through life-cycles of deployment and maintenance and to support robust provenance of data arising from these sensors (Sarbhai et al. 2019). Moreover, implementation of blockchain technology will allow us to provide much higher control of data access.

16.5.2 Participant-Facing Tools

Exposure studies involve multiple stakeholders: participants and their families, clinical coordinators, researchers, sensor developers, and system administrators. In order to meet the needs of all these stakeholders, we utilized user-centered design (UCD) approaches (McMullen et al. 2011) to develop methods for their interactions with EHIE, including data collection, visualization, and analysis. UCD is a multidisciplinary approach rooted in cognitive and behavioral science, based on deep understanding of who will be using the system, their tasks, expectations, and contexts of use. UCD methodologies are congruent with the international standard ISO 9241-210:2010, Ergonomics of Human-System Interaction. In this section, we cover the tools and methods for participant interactions with EHIE.

We use participant-facing tools in exposure health studies to collect participant-reported data prior to the start of a study and during the study phases. We also use these tools to support the collection of data representing participant behavior and environmental sensing, as well as provide feedback and inform participants about interventions for different study designs (Figure 16.8). These interactions could be triggered by environmental and physiological sensor-, clinical-, or participant-reported events.

We developed processes and methods for selection, use, and integration of various types of tools for different study designs. A video titled "Utah PRISMS Informatics Ecosystem" shows these tools in use and can be viewed at https://www.youtube.com/watch?v=FT7Yz5l94fQ. For demonstrative purposes, we used exemplar tools including a generic clinical study tool, a domain-specific exposure health study tool, a visualization tool, and annotation tools:

1. *REDCap (Harris et al. 2009)*: A highly used, open-source study data management tool for designing and administering surveys and data collection. It is a Health Insurance Portability and Accountability Act (HIPAA)–compliant online platform. Participants use REDCap surveys in a PRISMS pilot study for providing their demographics and periodic symptoms. We integrate project-specific data from REDCap using its application programming interface (API).
2. *eAsthmaTracker (eAT) (Nkoy et al. 2012, 2013)*: Integrated asthma patient self-management, education, research and clinician communication platform with alerts developed by the Department of Pediatrics, University of Utah. It collects data on pediatric asthma symptomology through validated instruments to calculate daily and weekly asthma control scores. It also collects participant perceived exposures to asthma triggers (e.g., pollen). In addition, eAT presents participants with air quality indices,

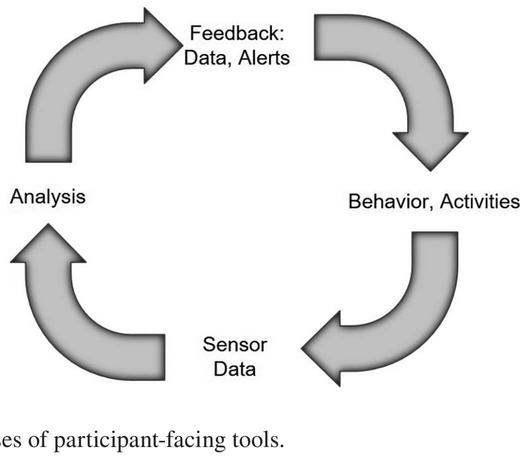

FIGURE 16.8 Uses of participant-facing tools.

pollutant levels from the Environmental Protection Agency (EPA), weather, and pollen counts. We integrate participant symptoms, daily and weekly asthma scores by accessing eAT's MySQL database.

3. *Visualization (Moore et al. 2018)*: We use open-source Grafana software (Grafana 2019) to provide real-time visuals of sensor data to participants in their homes.
4. *Activity annotator (Moore et al. 2018)*: Often, sensor readings require additional information on activities performed by participants or happenings in their environments to provide context to data and its analysis. The activity annotator allows participants to make free-form or prompted annotations, linked to artifacts on the sensor data feed. For example, our participants would note activities such as cooking or cleaning which caused a spike in particulate counts. We examined three approaches to annotation: text messaging, visualization screens (see above), and voice annotations using Google Home. We stored these annotation data into a MongoDB-based annotation document store that then can be consumed for data integration.
5. *Ecological momentary assessments (EMAs) (Shiffman, Stone, and Hufford 2008)*: Obtain participant behaviors and experiences in real time and in their natural environments using short answers to questions. We used EpiFi along with REDCap and Twilio (Twilio 2019) to administer EMAs as text messages that were triggered randomly, periodically at specific dates and times, or by sensor measurement artifacts (e.g., assessments triggered by spikes in sensor readings). See Figure 16.9. Participant responses via text messages were then recorded into REDCap and made available for study data integration, as described above.

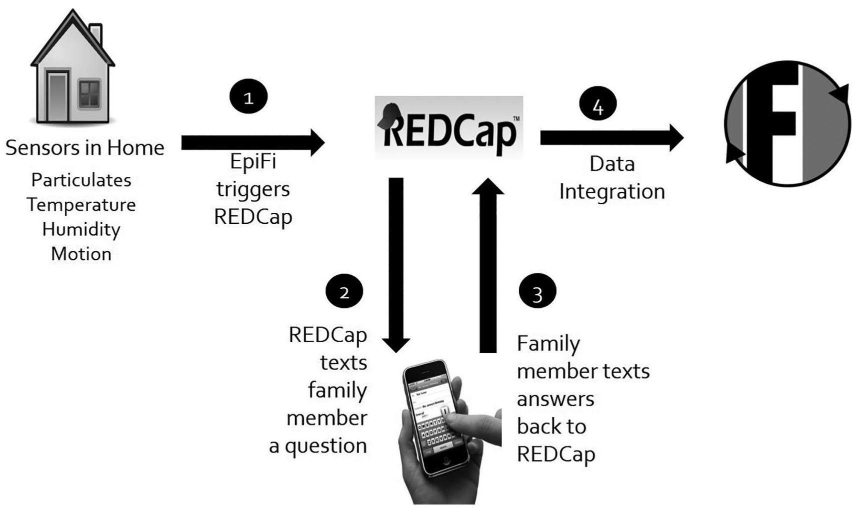

FIGURE 16.9 Ecological momentary assessments administered using text messages and orchestrated by EpiFi, REDCap, and Twilio.

16.5.3 Researcher-Facing Tools

Exposure health studies vary widely in design and purpose and, hence, researcher-facing tools must support a wide spectrum of requirements. The studies may range from sensor development projects to highly complex epidemiologic, retrospective observational, or prospective observational designs. Additionally, many of these studies entail real-time interactive data collection (e.g., EMAs) (Gouripeddi et al. 2017, Habre et al. 2016). Both primary and secondary sensor data may be used in studies. The use of primary sensors necessitates collection, storage, and management of real-time data, often with substantial data volume. The sensors may be personal and wearable or stationary, positioned indoors or outdoors, and may measure a wide variety of variables associated with multiple species. If measurements or study interventions must be triggered by sensor values, additional decision support tools are needed to support or automatically initiate the appropriate actions. Secondary sensor data may or may not be pre-processed prior to use, can be high in volume, and details about any pre-processing may be elusive.

The sensor data collected in these studies must often be analyzed jointly with a similarly diverse variety of variables and data types, including patient-reported outcomes, activity annotations, self-reported symptoms, electronic health record data, biospecimen data, computational models, and aggregates of these data types. For this wide variety of data that is used in diverse ways, researchers require support for the collection, monitoring, and integration of data. They also require tools that help them design and monitor studies, with alerting when participants or sensors require attention. In order to achieve the goals of the study, researchers also need additional tools that aid analysis.

Using UCD has allowed us to use some of the same tools used for participants to provide researchers with tools for system interaction. Here, we present examples of researcher-facing tools corresponding to different stages in the life-cycle of an exposomic study:

1. *Study design*: We developed and use process diagram templates (Figure 16.10) for describing study recruitment, sensor deployment, and data integration pipelines that a researcher could design and then can be used by the research administration team to assemble various artifacts.
2. *Sensor selection*: Exposure health studies often use sensors to measure participant environments and physiology. In order to allow researchers to select appropriate sensors, we developed a library that stores detailed descriptions about sensors, which researchers can browse in order to select appropriate sensors for their study needs (Burnett et al. 2018b). The sensor library (Figure 16.11) includes metadata about the owners of the device, methodology used to measure environmental and physiological species, device characteristics (e.g., battery usage), calibration, validation and measurement details, as well as an inventory of number of available devices. We deployed the sensor library using a Neo4j graph database (Neo4j 2019). We are similarly developing a library of computational models.

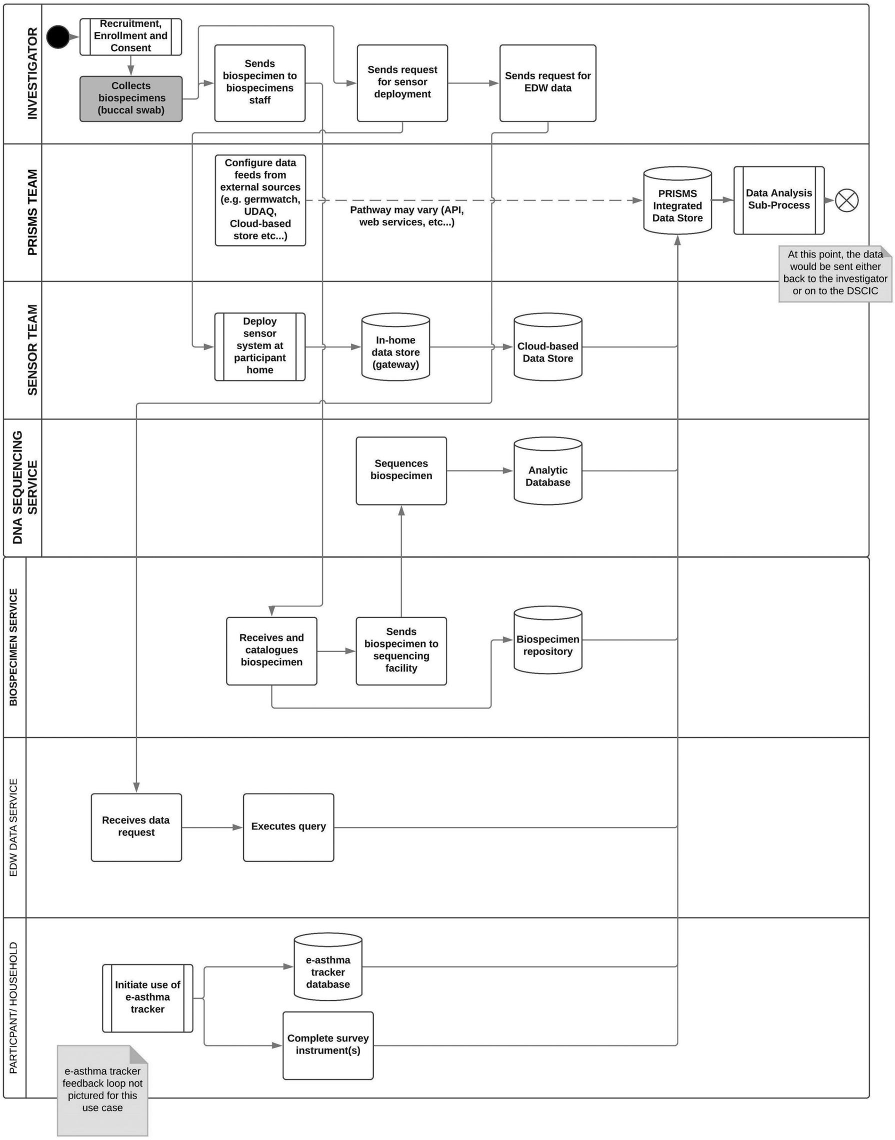

FIGURE 16.10 A high-level process diagram of a pilot study used to evaluate EHIE depicting data streams and their management, including their collection, integration, and final submission.

3. *Data and metadata collection*: At times, the research team itself would need to collect study data, metadata, or other information when performing studies. Using REDCap projects, we can collect:
   a. *Dwelling unit survey (Jacobs et al. 2009)*: Details about home environment including structural characteristics and indoor features such as

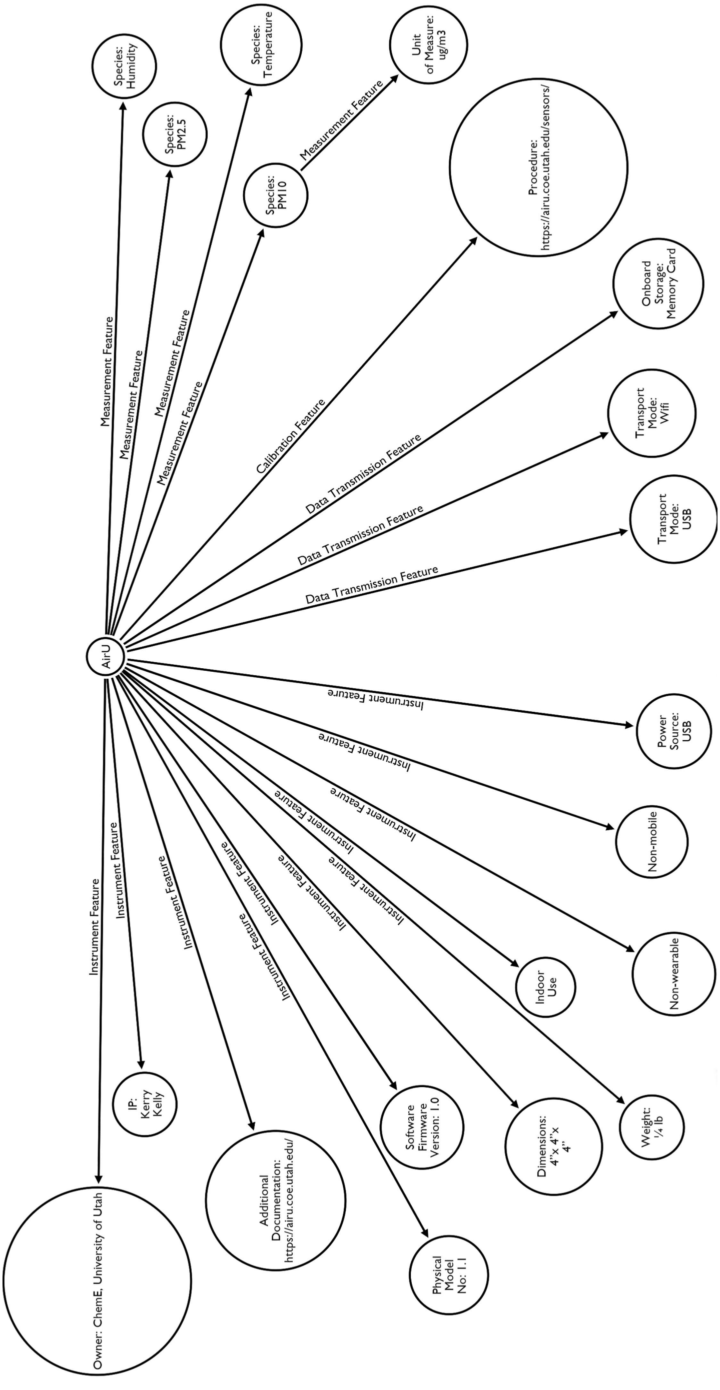

FIGURE 16.11 A section of the metadata about the AirU sensor (Kelly et al. 2017) stored in the sensor library.

carpeting, appliances, heating, ventilation, and air conditions, which can be used for study analysis. Sensor deployment metadata: information about which sensors and where/when they have been deployed is then ingested into the DMDR (i.e., a MongoDB store) (Figure 16.5) for use within the acquisition and integration pipelines.

b. *Sensor metadata collection (Burnett et al. 2018b)*: Collection of detailed metadata describing sensors (e.g., measurement species, unit of measure, detection limit, sample volume, temperature, humidity) by means of an easy-to-use online tool that can be filled by the sensor owner and then ingested into the Neo4j store of the sensor library. Forms for collecting these metadata are available at http://j.mp/2U3Ixqw.

4. *Study monitoring*: We provide visualizations using the Grafana platform and the deployment status page to get real-time health of each deployed sensor.
5. *Study data analysis*: Various data can be integrated by the central Big Data integration platform. These data are available in different analytical data models or custom formatted databases or as exports. These data can be analyzed by the research team using traditional statistical software (e.g., SAS) or machine learning packages (e.g., R, TensorFlow (Abadi et al. 2016)). They can also be consumed in analytical pipelines as shown in Figure 16.12 or through dashboarding tools.

16.5.4 Computational Modeling Platform

It is often not possible to measure participant environments with resolution necessary to perform exposure health studies. Also, participants are not always in the proximity of a sensor to measure their environmental exposure. For example, the sparse placement of the EPA air quality monitoring stations in Salt Lake County fails to capture fine-grained variation in air quality due to weather, elevation, and topography. Basic interpolation methods are insufficient for capturing this variation,

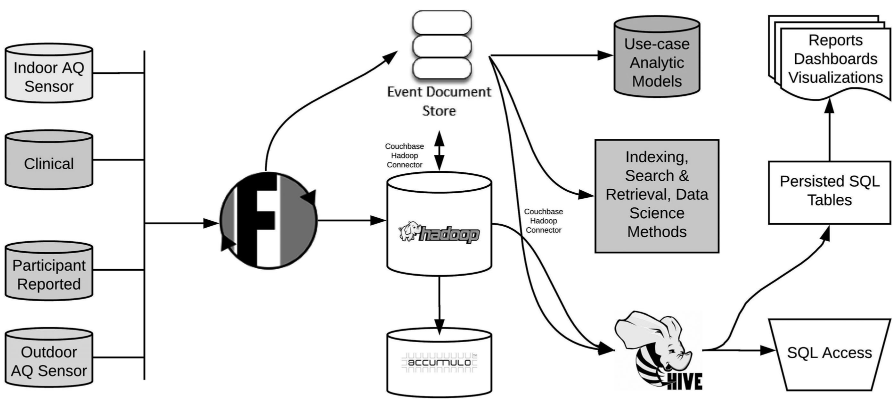

FIGURE 16.12 An example analytic pipeline used in EHIE.

due to geographical and structural irregularities. Increased monitoring (e.g., fine grid or mobile monitoring) is expensive to implement. These limitations can be overcome using computational models to fill gaps in the measured data. In this section, we describe developments within EHIE related to computational modeling:

1. *Activity and location recognition*: While there are multiple models to predict and simulate environmental measurements, a major gap in using these models for exposure health studies is the assignment of either measured or modeled levels of environmental species to participant locations at different times. Human populations are mobile. In addition, situational context of what activities individuals are performing is also important. While it is possible to track locations and activities of participants, there are several major issues, for example, GPS devices do not work reliably in indoor settings, privacy issues complicate activity tracking, and population-scale monitoring is difficult and expensive. To address this, we developed the Spatio-Temporal Human Activity Model (STHAM), which assigns locations and activities to individuals based on their demographic profile. This model can then be integrated with exposure levels to generate comprehensive spatio-temporal records of exposure (Figure 16.13).

   STHAM is an agent-based Monte-Carlo model (Sward et al. 2017, Lund, Gouripeddi, and Facelli 2019a,b, Lund et al. 2017). It is semi-empirical and utilizes externally collected datasets: (1) 2010 US Census (US Census

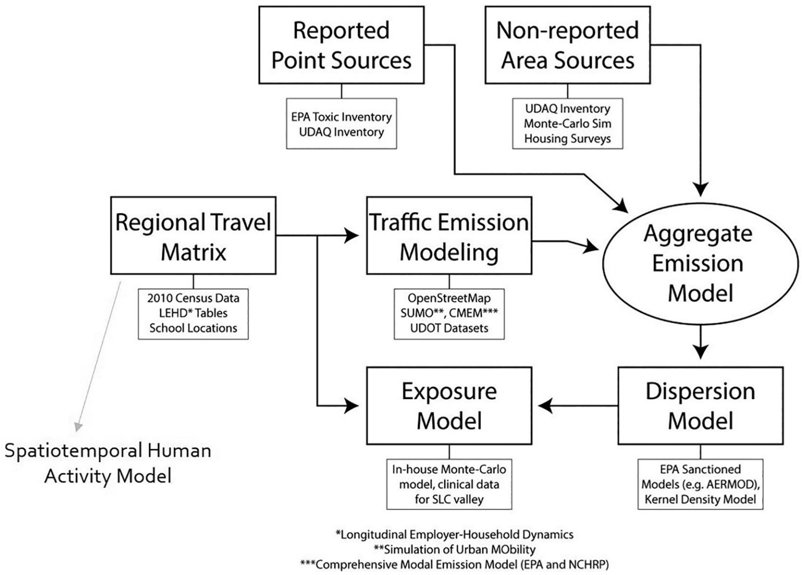

FIGURE 16.13 Overview of the Spatio-Temporal Human Activity Model (STHAM), which integrates simulated activities and locations to develop comprehensive spatio-temporal records of exposures.

Bureau 2010), (2) American Community Survey (US Census Bureau 2017), (3) US Bureau of Labor Statistics American Time Use Survey (ATUS) (U.S. Bureau of Labor Statistics 2017), and (4) Census Bureau Longitudinal Employer-Household Dynamics (LEHD) Origin-Destination Employment Statistics (LODES) (US Census Bureau Center for Economic Studies 2016). Steps taken in building STHAM are summarized in Figure 16.14.

We briefly discuss the use of STHAM for the Salt Lake, Utah, metropolitan area:

a. *Household assignment*: Using data from the census, we assign an age category (0–17, 18–64, 65+ years old) for each household in Salt Lake Valley, based on age of majority of members in that household.
b. *Activity classification*: The ATUS dataset has approximately 10,000 respondents each year and includes their daily activity diaries consisting of about 500 macro activity categories, demographic data, and contextual information for each activity. We performed unsupervised classification using random forests, followed by a dimensionality reduction using t-distributed stochastic neighbor embedding (t-SNE) and, finally, a density-based clustering to result in 90 demographic classes and 40 activity day classes.
c. *Activity sequence construction*: Humans usually follow a schedule, but typically have intra- and inter-person variations. To account for this, we used the activity classes to construct activity windows, probabilistically sorted these windows, and then used a Monte-Carlo based activity generator to construct sequences of activities.
d. *Assessing diurnal patterns*: In the final step, we probabilistically assigned demographic classes to a day type and built activity sequences for each demographic class based on day type. We then assigned these demographic-specific activity sequences at an hourly basis to population distributions at a 500-m grid level.

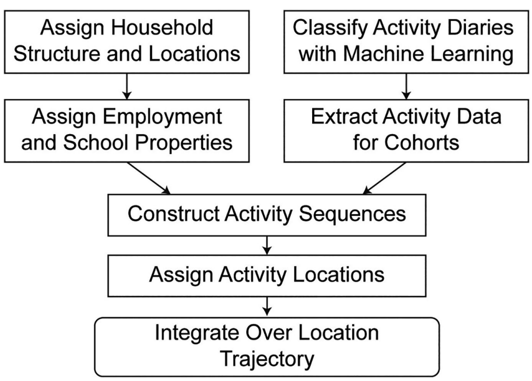

FIGURE 16.14 Schema of steps used in building STHAM.

2. *Indoor activity recognition*: Sensor readings on their own might not always be sufficient to get contexts of triggers leading to changes in environments. We, therefore, have developed methods for participants to annotate sensor data using various approaches as described above. But, often participants are overburdened annotating their activities. In order to automate some of this, we trained machine learning methods with existing annotations and sensor streams to predict activity signatures from the profile of the sensor readings.
3. *Personal exposure modeling (An et al. 2018)*: Estimating exposures at a personal level not only needs to account for the environment, locations, and activities of individuals but also their biological characteristics. We extend our current work from STHAM to such multi-scale models that can then be utilized in research studies.
4. *Modeling metadata of computational models (Lund et al. 2018)*: Many environmental modeling algorithms have been developed by the EPA and other groups (e.g., US Environmental Protection Agency 2016a–c, McMillan et al. 2010, US Environmental Protection Agency 2005, Yanosky et al. 2014). We are also able to accommodate models of different resolutions, such as those available from Mesowest (Horel et al. 2002). All these models have varied capabilities, methods, limitations, deployment needs, inputs and simulated output details. In order to meet the needs of varied exposure health studies, we developed a library of modeling algorithms that can be utilized to generated high spatio-temporal resolution data. This library captures technical details and methodology of each model, their deployment characteristics when used in a particular study, and describes inputs consumed and output generated from these models. Similar to the sensor library, this library of computational models supports selection and use and points to software code that can be executed by workflow engines (Deelman et al. 2015). We implemented the library using a Neo4j graph database platform, along with detailed metadata about each modeling algorithm.
5. *Uncertainty quantification (Gouripeddi et al. 2015, Burnett et al. 2015)*: The computational modeling platform also includes methods to quantify different types of uncertainties that might be present within data or associated with computational models. Uncertainty could be: (1) Inherent: Variations in unknown conditions, (2) Reducible: Associated with the model and input conditions, and (3) Exposure-related: Arising due to differences in person's exposure and true ambient environmental levels (Zeger et al. 2000). For example, STHAM can only provide ranges of activities and expected locations, but not actual behaviors. However, in this example, the model could be validated by proxy, by evaluating simulated traveling numbers against actual measured traffic volumes. Activity or context measurement creates a new and complicated dimension of uncertainty. Multi-agent contributions to activity context also need to be measured or simulated (e.g., two people in the same household contribute to and experience the activity context differently).

16.5.5 Central Big Data Integration Platform

A key challenge to performing exposomic research is integrating multiple sources of data for generating comprehensive records of exposure. Such an integration of data needs to support the following features (Sward et al. 2017, Gouripeddi et al. 2019a):

1. Generate comprehensive spatio-temporal records of exposures.
2. Semantically consistent metadata-driven integration of heterogeneous data.
3. Maintain spatio-temporal integrity and support reasoning.
4. Associate uncertainties with using data as exact quantifications of exposure.
5. Transform data to support diverse translational research archetypes.

EHIE leverages and extends the OpenFurther (OF) data integration and federation platform (Bradshaw et al. 2009, Livne, Schultz, and Narus 2011, Gouripeddi 2019, Gouripeddi et al. 2012, 2013). Main components of OF include an Ontology/terminology Server (OS); a Metadata Repository (MDR); SS, which can be consumed by various tools; Data Source Adapters (DSA); Administrative and Security Components (ASC); Virtual Identity Resolution on the GO (VIRGO); Quality and Analytics Framework (QAF); a Computational Modeling (CM) and Uncertainty Module; Process-Workflow Module (PWM); a Metadata & Semantics—Discovery and Mapping Service; a Knowledge Repository; and a Federated Query Engine (FQE) that orchestrates queries between the PWM, SS, MDR, OS, DSA, ASC, VIRGO, CM, and QAF. For EHIE, we modified OF with an Event Document Store (EDS) that stores integrated events as events in a Big Data store; and used graphical stores for the MDR. OF supports the selection and integration of heterogeneous data for generating high-resolution spatio-temporal grids of exposures and associates this data with characterized metadata to support their proper utilization. By leveraging characterized metadata and semantic mappings that are stored within a MDR and an OS, respectively, OF provides syntactic and semantic interoperability for dynamically federating data and information. This federation can take place in real time or statically, without requiring data owners to extract and/or transform their data—facilitating integration by retaining data in their native format. Using this approach, OF is able to transform and integrate distributed data across multiple scales, models, and semantics into consumable formats.

In order to meet the above-mentioned requirements for exposomic research, we modified OF as follows:

1. Sensor Common Metadata Specification (SCMS) (Burnett et al. 2017)
   Quantifying exposures and their effects requires integration of multiple sensors that measure the general and personal environments for chemical, physical, and biological species. Even within a given species, there are often differences in their composition profiles based on their source and locations (Yang et al. 2019). Sensors used to measure these species have different instrument characteristics, capabilities, calibrations, and outputs (Williams et al. 2019). Integration of diverse sensor data should be context-aware,

metadata-driven, and semantically consistent. Further, it needs to be supplemented with metadata to support appropriate use of data as well as provide a harmonized representation for ease of use in diverse research studies and analytic approaches. In order to support these needs, we developed SCMS by performing a literature review of studies using sensors, reviewing sample data, and iteratively refining it with feedback from sensor experts. SCMS is available at https://github.com/uofu-ccts/prisms-sensor-model for community review and utilization.

The scope of SCMS includes all types of sensors ranging from nanosensors to satellites, measuring physical, chemical, or biological species. These sensors could be personal or mobile, stationary in-home, or ambient monitoring stations. SCMS covers three sensor domains:

- *Instrument*: Physical characteristics of a sensor device.
- *Deployment*: Description of how a sensor device is used in research data collection.
- *Output*: Characteristics of sensor measurements.

The contents of these three domains ensure quality in exposure studies by providing the content and structure for (1) establishing a library of sensors as described in Section 16.5.3, (2) development of a DMDR also described in Section 16.5.1, and (3) the development of data harmonization MDR (described next). Using SCMS, we develop a similar metadata model for interventional devices used in research studies (Morgan, Gouripeddi, and Sward 2019).

2. Metadata management (Gouripeddi et al. 2019b,c, Sward et al. 2017).
OF's MDR (Bradshaw et al. 2009, Mo et al. 2014) is an Object Modeling Group specification conformant (Object Management Group 2019), FAIR-compliant (Wilkinson et al. 2016), standard-based repository of artifacts and knowledge. It stores metadata artifacts and relationships of data and modular components subscribed by OF. These artifacts include, but are not limited to: (1) logical models, local models, model mappings, (2) administrative information, (3) descriptive information, and (4) translation programs. These are organized as "assets" in a custom-built, highly generic and abstracted entity-relationship model. Assets may have properties and associations to other assets. Stored metadata is shared in various structured and non-proprietary formats using translation programs and made available for consumption by different SS.

Considering the data and process complexity within exposomic research, we conceptually divided metadata management into three categories:

- *Data metadata*: Metadata that describes data outputs resulting from an observations or measurements. This includes sensor measurements, outputs of computational models, clinical observations, genomic sequence annotations, socio-behavioral data, and participant report data.
- *Process metadata*: Metadata that describes research or data processes within EHIE. This includes sequences of steps followed in different computational models in order to generate outputs, and data

transformation and integration workflows to harmonize source data as events or into analytical models. An example of a research process is sensor deployment.

- *Knowledge resources*: Metadata that describes a source or instrument used to collect, measure, or derive data. This includes sensor devices, electronic medical records, or study-specific data collection instruments.

The SCMS has highly interconnected metadata elements. The labeled property graph provides better support for (1) complex relationships, such as ternary or higher degrees, many-to-many, and self-referencing relationship types, and (2) dynamic schemas (Robinson et al. 2015). Unlike relational stores, graph metadata management does not require deviating from natural relationships representing semantically rich domains in the real world. We, therefore, adopted a graph-based MDR for the instrument and output domains of SCMS which represent data metadata and knowledge resources aspects of metadata categories. Since the deployment domain of the SCMS is fairly simple with one-to-many relationships, we adapted a document store that captures metadata about the deployment processes as described above. We implemented the instrument domain in a Neo4j graph database (Neo4j Graph Platform 2019), which forms the backend store for the sensor library (Burnett et al. 2018a), as described in the participant-facing tools section. OrientDB graph database system (OrientDB 2019) provides better integration with Java classes. We, therefore, used OrientDB for storing data metadata which includes data transformation functions to transform source data into events. Similarly, we have adapted OrientDB for storing data metadata of other data domains such as clinical, biospecimen, socio-behavioral, and participant report which we were originally stored in relational databases. We adapted the MongoDB platform (MongoDB 2019) for the deployment MDR as described in Section 16.5.1.

These new approaches to metadata management limit the introduction of semantic dissonance between conceptual models and their implementation, as it retains complex real-world relationships. Detailed metadata stored in the MDRs provide a spatio-temporal grid of metadata complementing the high-resolution spatio-temporal grid of integrated data, informing end-users of the limitations and uncertainties associated with the data. As a next step, we plan to characterize, develop specifications, consume, and store relevant semantics for the exposome domain (Habre et al. 2016, Mattingly et al. 2016, Burnett et al. 2018a, Kiossoglou et al. 2017, Gouripeddi, Habre, and PRISMS Data Modeling Working 2019, Cummins et al. 2019a, Lopez-Campos et al. 2019) and supplement this metadata work.

3. Event Document Store (EDS)

Exposomic studies require integrity of spatial and temporal dimensions of data in order to ascertain relationships between patient-reported symptoms, physiological measurements, and clinical manifestations, with environmental changes (Gouripeddi et al. 2017). This requirement for generating spatio-temporal records of exposures and the need to provide data in different analytic models and formats led us to transform all data as events occurring in spatio-temporal coordinates. Based on the uncertainty associated with the proximity of data collection to the subject under consideration, and the usage of different data collections in different types of translational research, we classified events into six domains: sensor, clinical, biospecimen-derived, patient-reported, computationally modeled, and aggregates (Figure 16.15) (Gouripeddi et al. 2017). These events are informed by logical models stored in the MDR. Each event is logically a document and we call the aggregation of all these documents an EDS. The EDS is conceptually modeled as events occurring on a timeline and can be implemented in any Big Data store. This primitive storage format allows linkage across different root objects that do not necessarily belong to a person and can be transformed into higher/analytical models based on use cases. Time represented in these events is modeled as (Combi, Keravnou-Papailiou, and Shahar 2010):

- *Unbounded*: Contains upper and/or lower bounds with respect to its order relationship.
- *Dense*: An infinite set of smaller units.
- *Discrete*: Every element has both an immediate successor and an immediate predecessor, if unbounded, and within the bounds, if bounded.
- Instants and intervals (upper and lower time points).
- Finest granularity available with the source.

Similarly, spatial dimensions in these events are continuous and transformable to different reference systems. We stored these events as JavaScript

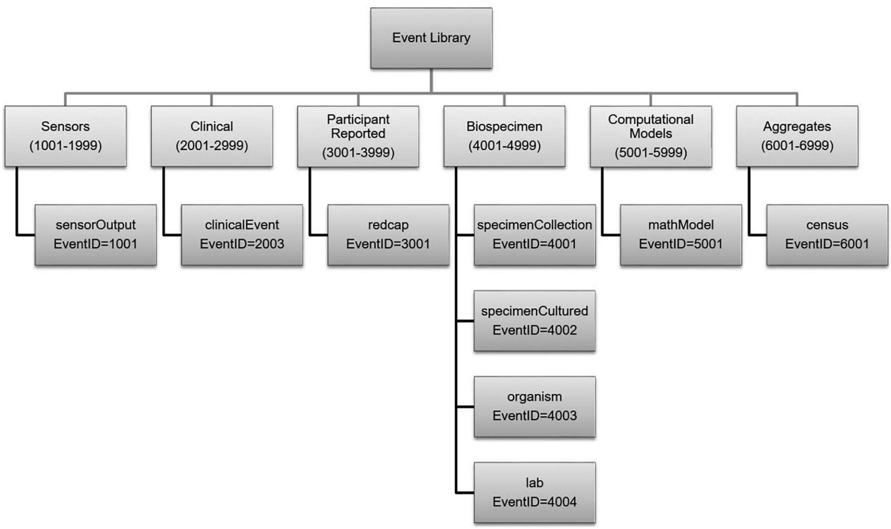

FIGURE 16.15 Conceptual representation of different events implemented with allocated bins of event identifiers.

object notation (JSON) documents in the Couchbase NoSQL platform (Couchbase 2019). The EDS supports (1) natural querying for spatio-temporal reasoning of events and knowledge facts—a critical need for complex and unpredictable diseases in which sequences and locations of events are critical to their understanding and (2) transformation of events into higher/analytical models to support diverse translational research archetypes.

4. Software workflow

OF transforms and stores data from heterogeneous sensors and other health data sources into uniform event-based data structures. In order to facilitate these data transformations, we modified OF to be an event-driven architecture with the following changes (Figure 16.16): First, data services identify user input data criteria for integration. Then, using the contents of the MDR, OF orchestrates querying of data sources which could be web services, database tables and flat files for attributes described in the user's input. OF's SS leverage metadata content in the MDR to inform structural and semantic transformations of selected data to their corresponding events. Using this metadata, OF's services then write events into the EDS. For example, home-based sensor measurements acquired via EpiFi (Lundrigan et al. 2017) are transformed into sensor events and integrated as JSON documents in EDS (Figure 16.15). OF also exposes several services to access and view the documents stored in the EDS.

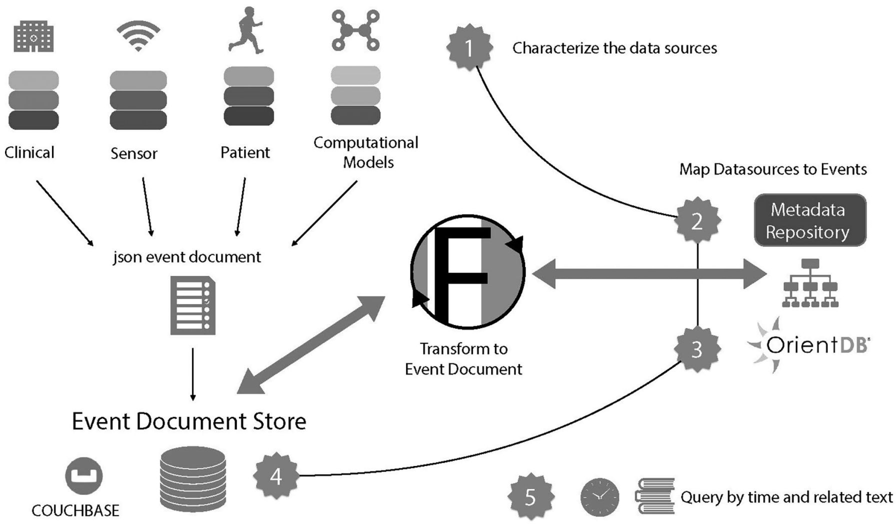

FIGURE 16.16 OpenFurther (OF) software workflow for exposure health studies. Key steps: (1) data sources are characterized, (2) their metadata and mappings to different event types are authored, (3) metadata and transformation functions are stored in an OrientDB-based graph MDR, which is (4) leveraged by OF to generate events stored in a Couchbase event document store, and (5) these events are available for querying by time and other downstream analytic processes.

Data in EDS is stored on a study by study basis. A research study is represented by a study event and consists of one or more integration events that represent periodic data integration runs. We found that the architecture can be scaled for performance. The architectural designs of the infrastructure support a semantically consistent, metadata-driven approach to multi-scale, multi-omics exposomic Big Data integration for diverse translational research ranging from understanding mechanisms of disease to developing preventive and therapeutic interventions. The developed architecture alleviates informatics challenges associated with exposomic data that originate from the characteristics of measurement devices, their deployment, and human behavior.

We index the data using Couchbase's native indexing and with ElasticSearch (Kuc and Rogozinski 2013) to support different spatio-temporal reasoning use cases. In addition, we use a similar metadata-driven approach described above to transform events into higher analytical models depending on the use case, which is then made available via different research-facing tools. We have also submitted the integrated data through data streaming pipelines such as the Kafka (Narkhede, Shapira, and Palino 2017) instance hosted at the PRISMS data coordinating center (Stripelis et al. 2017).

16.6 UTILIZATION OF EXPOSOME IN TRANSLATIONAL STUDIES

We are utilizing EHIE to generate exposomes for two ongoing studies:

1. *PRISMS Pilot (Gouripeddi et al. 2019b,c)*: This study was approved by the University of Utah Institutional Review Board (IRB No._00086107). We adopted an ongoing transient receptor potential pediatric asthma study (Deering-Rice et al. 2015, 2016), where we collected and integrated environmental data with health data for 10 participants residing in Salt Lake, Davis and Utah Counties, Utah, United States, for the period March 1st, 2017 to June 30th, 2018 (study processes depicted in Figure 16.10). A summary of the data used in this study is presented in Table 16.1. We integrated this data into approximately 25 million events for the study period (Figure 16.17). The generated events included participant registration, clinical, survey, sensor output, and sensor deployment events, linked by events representing different integration batches and the study. On evaluation of the quality of the integrated documents, we found the events to be consistent and accurate with the source data. In addition, we were able to perform analysis of these events to ascertain spatio-temporal relationships between various events. We submitted this information to the PRISMS data coordinating center in order to test the PRISMS program concept. Statistical and machine learning analysis results being performed both at the University of Utah and the data coordinating center are pending.
2. *Environmental influences on Child Health Outcomes (ECHO) Study (Collingwood et al. 2018)*: This study was approved by the University of Utah Institutional Review Board (IRB No._00086107). As part of ECHO, we piloted deployment of sensors among urban, rural, frontier, and tribal populations to evaluate the acceptability of low-cost, IoT connected air

TABLE 16.1
Summary of Data Integrated as Events for the Utah PRISMS Pilot Study for the Period from March 1, 2017 to June 30, 2018

Data Stream	Description	Event Count
Participant demographics	Collected using a REDCap questionnaire. Data consumed via the REDCap API (Harris et al. 2009)	10
Participant home assessment	Collected using a REDCap questionnaire. Data consumed via the REDCap API	10
Home sensor deployment	Details about the indoor home sensor deployment. Collected using a REDCap questionnaire. Data consumed via the REDCap API	145
Asthma severity assessment (weekly)	Weekly asthma symptoms and severity scores of participants collected from eAsthmaTracker (eAT) (Nkoy et al. 2012, 2013). Data consumed from MySQL database of eAT	293
Asthma severity assessment (daily)	Daily asthma symptoms and severity scores of participants collected from eAT. Data consumed from MySQL database of eAT	1,160
Particulate matter (PM2.5) for Salt Lake, Davis and Utah Counties, Utah, United States	Environmental Protection Agency's Air Quality Datamart API ("AQS Data Mart | Air Quality System | US EPA" 2019)	73,995
Temperature, humidity for Salt Lake, Davis and Utah Counties, Utah, United States	National Weather Service API (US Department of Commerce 2019)	703,164
Particulate matter (PM2.5) from PurpleAir	Data from Purple Citizen's Network (PurpleAir.Org 2019) aggregated by Mesowest (Horel et al. 2002)	1,738,132
Particulate matter (PM2.5) from Trax light rail in Wasatch front	Mobile Air Quality Assessment from Trax (Mitchell et al. 2015, 2018) aggregated by Mesowest	7,934,529
Indoor particulate matter (PM2.5) measured by Utah Modified Dylos	Data made available through EpiFi (Lundrigan et al. 2017) and consumed from an influx database	14,075,201
Total		**24,526,639**

quality measurement devices varied populations. With good results in our pilot testing, we deployed multiple sensors in homes of participants recruited via a geographic random sampling protocol whereby the primary inclusion criteria was a women of childbearing age residing in the home. To date, we have deployed multiple sensors in 28 homes of participants and collected in excess of 50,000,00 sensor readings measuring PM2.5 using AirU devices (Kelly et al. 2017) through EHIE over a 1-year period.

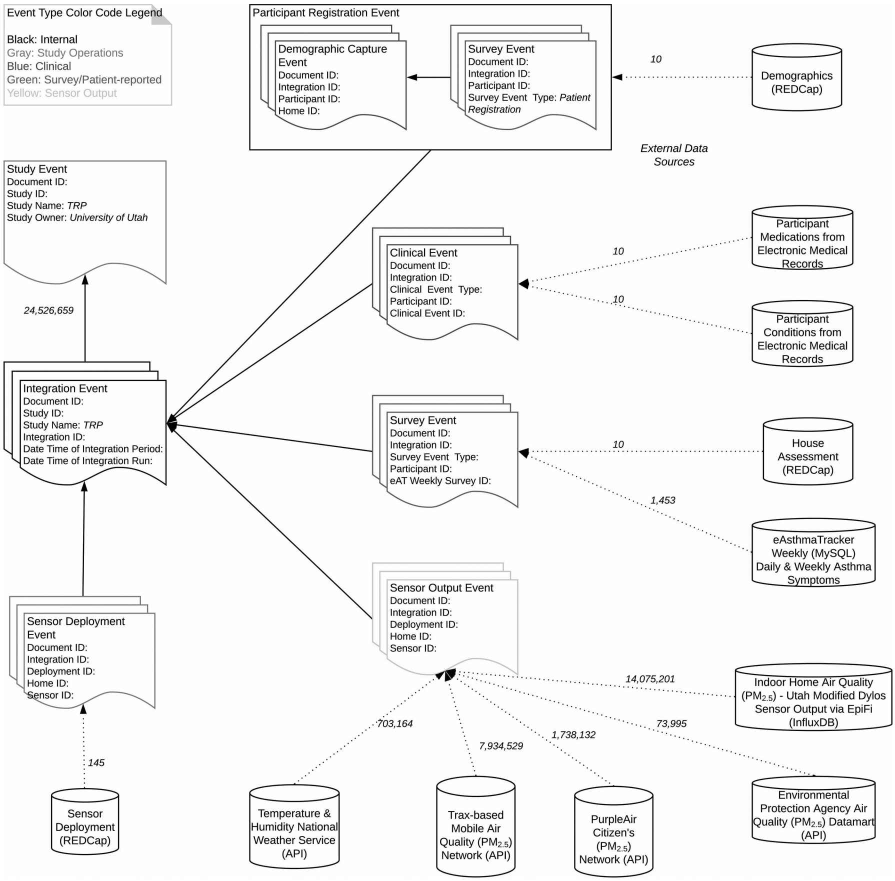

FIGURE 16.17 Different types of events generated by integrating sensor and clinical datasets in the Utah PRISMS pilot study. Data from ten participants for the period March 1st 2017 to June 30th, 2018 and residing in Salt Lake, Davis and Utah Counties, Utah, United States, were integrated resulting in a total of 24,526,659 events. Source data consisting of participant registrant data (bottom right to left) including participant demographics and criteria of their eligibility to the study, clinical data, home assessment surveys, asthma symptoms, indoor and outdoor air quality sensor readings, weather and detailed sensor deployment data. These events are linked via integrated events and study events. Counts of events integrated from each source are indicated on top of arrows connecting them to their corresponding event types.

Ongoing investigations relative to these environmental sensors include in-home air quality variation, relationship between indoor/outdoor & housing characteristics, individual sensor measurement variation (drift), and home-based air quality influences on urinary markers of inflammation. Future work will expand the deployment to more than 200 AirU sensors and in 60 additional homes. In addition, the measurement capability of the AirU's

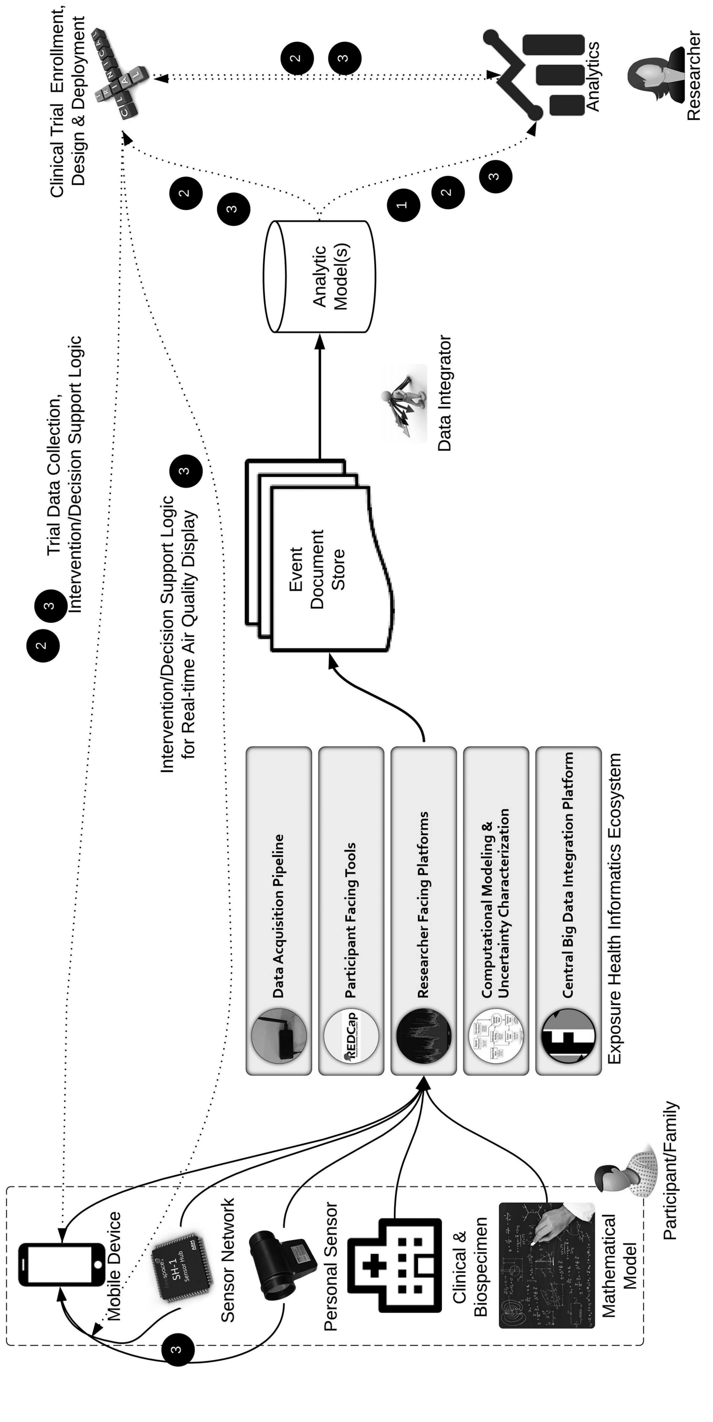

FIGURE 16.18 Diverse translational exposomic research use cases supported by EHIE leveraging IoT Devices and Big Data methods.

will be enhanced by embedding ozone and volatile organic compound (VOC) sensors into the unit, using EHIE for management of deployment processes, and integration of data arising from over 250 sensors.

Our initial results from these pilots show that it is feasible to use informatics ecosystems like EHIE to perform sensor-based longitudinal cohort studies (Collingwood et al. 2018, Gouripeddi et al. 2018, Sward 2019, Cummins et al. 2019b). While different sensors capture data at different time resolutions and have varying spatial distributions, it is possible to generate spatio-temporal grids of exposures and relate to symptoms and clinical observations. Most participants were supportive of using their personal WiFi for sensor data transmission, but there is a need for flexible data transmission and storage methods to account for limitations in home networks. Preliminary statistical analysis of the data using mixed modeling approaches shows that particular matter (PM) levels were generally lower indoors compared to outdoors ($p<0.001$). Indoor PM counts were not related to asthma control test (ACT) scores and rescue medication usage ($p>0.05$). Outdoor PM2.5 was related to worsened ACT scores and increases in asthma treatment ($p<0.001$), but outdoor PM10 counts were not related ($p = 0.52$).

16.7 DISCUSSION AND CONCLUSION

In this chapter, we discussed challenges and data needs for performing total exposure health research (Sward et al. 2017). These include (1) acquisition of sensor data, (2) selection of heterogeneous data sources, (3) filling gaps in measurements using computational modeling, (4) characterization of uncertainties associated with data, (5) generation of high-resolution spatio-temporal grids of exposures, (6) integration of data, (7) presentation and visualization of data, and (8) support for diverse set of translational research archetypes. In order to address these needs, we developed a loosely coupled, scalable informatics infrastructure called EHIE that consists of (1) data acquisition pipeline, (2) participant-facing tools, (3) researcher-facing platforms, (4) a computational modeling platform, and (5) a central Big Data federation/integration platform. For each of the above components, we provide an introduction and discuss their subcomponent architecture. The ECHO and PRISMS pilot are examples of demonstrative studies utilizing EHIE to generate exposomes for research.

EHIE is a generalizable, multi-scale, and multi-omics platform providing robust pipelines for reproducible exposomic research that uses real-time, low-cost sensors to provide spatio-temporal records of environmental exposures. Using this component-based ecosystem, we are able to support the deployment and performance of sensor-based studies, and the integration, processing, visualization, and secure transmission of study data for most research designs. EHIE provides an effective, flexible, and open access approach to collecting, managing, and analyzing high-resolution data from sensors. Because the infrastructure is based on logical data models for clinically relevant exposomes (environmental exposures such as air quality, which have health impacts), the infrastructure is flexible and adaptable to many translational research scenarios (Figure 16.18). The infrastructure also provides mechanisms for integrating exposure profiles (exposomes) with clinical, self-reported, behavioral, and other research data. In addition, EHIE supports supplementation of direct measurements with computationally-modeled data for exposures, activities and locations that can then easily be assimilated into comprehensive spatio-temporal records.

We have formalized the development and sustenance of this ecosystem and our team of collaborators as a Center of Excellence for Exposure Health Informatics (CEEHI) (http://ceehi.ccts.utah.edu/). CEEHI serves as a collaborative for continuing investigations and development of state-of-the-art informatics methods for exposomics. It is a go-to center for researchers interested in conducting sensor-based, mobile, and virtual studies that include measurements of the environment, physiology, and behavior of participants by providing expertise, guidance, infrastructure, and other resources to the total exposure health research community.

CEEHI provides a key infrastructure that accommodates diverse types of future studies, including general and personal environmental exposure monitoring, activities, and physiological responses to the environment. From an infrastructure perspective, CEEHI seeks to evaluate the ecosystemic health (Jansen 2014) of EHIE. We will then advance it as an ultra-large-scale infrastructure with integrated sensor health monitoring systems and with added abilities to perform studies using mobile sensors during real-life activity and location trajectories of participants (Bill Pollak 2006, Friedman et al. 2014). We will advance the use of novel sensors (Wang and Tao 2017, Li et al. 2019, Mirowsky et al. 2013), sensing paradigms (Schivo et al. 2013, McCartney et al. 2017, Hichwa and Davis 2018), and sensor networks and architectures (Kasera 2019). We will improve upon the science of using appropriate resolutions of data for different use cases by testing different data resolutions as available from the measured and modeled data available in networks such as Mesowest (Horel et al. 2002). We will add robust support for management of research processes and data for activities related to the study and its operations.

On a health research front, CEEHI is working with multiple researchers at the University of Utah and elsewhere to perform studies that seek to understand mechanistic, health outcomes, interventional aspects of exposure, and diverse disease conditions. Our current pediatric collaborators are interested in pediatric asthma and children with complex medical conditions in improving care, better self-tracking with environmental changes (Nkoy et al. 2012, 2013) and pharmacogenomics in relation to the environment (Deering-Rice et al. 2015, 2016). The Utah Children's Project, part of the ECHO Consortium (Stanford and Collingwood 2017), is continuing a longitudinal cohort study of children starting from pre-conception with plans to follow them through 20 years of age while measuring a broad array of environmental exposures. These include examining effects of diverse exposures (chronic and intermittent) on health and human development; investigating basic mechanisms and gene-environment interactions of developmental disorders and environmental factors (both risk and protective) that influence health and developmental processes; indoor and outdoor air sampling; home evaluations; questionnaires; and a diverse array of biomarkers including microbiome, serum antibodies, and others. In addition, components of the data acquisition platform are being used to investigate exposures in occupational environments (industry and military organizations), as well as environmental exposures of underserved populations (frontier and rural families). Other total exposure health conditions being studied include adult pulmonary disease (Pirozzi et al. 2015, 2017), sleep apnea (Sundar, Daly, and Willis 2013, Zanobetti et al. 2010, Weinreich et al. 2015, Billings et al. 2019), chronic kidney disease (Bowe et al. 2018, Kaskel et al. 2014), diabetes and metabolic disorders (Riches et al. 2019), aging and neurological conditions. In addition, at a mechanistic level, we are planning studies that look at the interaction of the microbiome with the exposome in asthma (Gouripeddi 2019a), and the role of PM chemical constituents in pulmonary health (Kitt et al. 2019). We are working on quantifying the digital exposome (Lopez-Campos, Merolli, and Martin-Sanchez 2017) and relating it with effects on health. Lastly, we are expanding the integration and real-time assimilation capabilities of the platform to obtain objective measures of autonomic nervous system physiology to detect real-time status of different conditions, such as impaired awareness of hypoglycemia (Groat et al. 2019, Mehta et al. 2019) and neuropathic pain (Singleton et al. 2008, 2014), and use them to develop interventions for managing these conditions.

ACKNOWLEDGMENTS

This research is supported by the Utah PRISMS Informatics Center (NIH/NIBIB U54EB021973), the Utah Center for Clinical and Translational Science (NCATS UL1TR001067), and the Utah ECHO Program (NIH UG3OD023249). Content is the sole responsibility of authors and does not necessarily represent official views of the NIH. Computational resources were provided by the Utah Center for High Performance Computing, which has been partially funded by the NIH Shared Instrumentation Grant 1S10OD021644-01A1. The authors acknowledge the contributions of all the team members affiliated with the Utah PRISMS Informatics Center grant, collaborators from the NIH PRISMS program, and the Program for Air Quality, Health, and Society, University of Utah.

REFERENCES

Abadi, Martin, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. "TensorFlow: A system for large-scale machine learning." In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, 265–283, Savannah, GA. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi.

Adame, Toni, Albert Bel, Boris Bellalta, Jaume Barcelo, and Miquel Oliver. 2014. "IEEE 802.11AH: The WiFi approach for M2M communications." *IEEE Wireless Communications* 21 (6): 144–152. doi:10.1109/MWC.2014.7000982.

An, Soyoung, Albert Lund, Nicole Burnett, Le-Thuy Tran, Peter Mo, Randy Madsen, Katherine Sward, R.K. Gouripeddi, and Julio Facelli. 2018. "Utah PRISMS ecosystem: An infrastructure for personal exposomic research." In *Ygnite* 2018, Washington, DC. http://yg.ksea.org/faq.html.

"AQS Data Mart | Air Quality System | US EPA." 2019. https://aqs.epa.gov/aqsweb/documents/data_mart_welcome.html.

Bakian, Amanda V., Rebekah S. Huber, Hilary Coon, Douglas Gray, Phillip Wilson, William M. McMahon, and Perry F. Renshaw. 2015. "Acute Air Pollution Exposure and Risk of Suicide Completion." *American Journal of Epidemiology* 181 (5): 295–303. doi:10.1093/aje/kwu341.

Iyer, Bala. 2014. "Understanding Software Ecosystems." Business, April 4. https://www.slideshare.net/balaiyer/software-ecosystem-visualization-methodology.

Boyle, Barksdale, Elizabeth, Nicole C. Deziel, Bonny L. Specker, Scott Collingwood, Clifford P. Weisel, David J. Wright, and Michael Dellarco. 2015. "Feasibility and informative value of environmental sample collection in the National Children's Vanguard Study." *Environmental Research* 140 (July): 345–353. doi:10.1016/j.envres.2015.04.006.

Billings, Martha E., Diane Gold, Adam Szpiro, Carrie P. Aaron, Neal Jorgensen, Amanda Gassett, Peter J. Leary, Joel D. Kaufman, and Susan R. Redline. 2019. "The association of ambient air pollution with sleep apnea: The multi-ethnic study of atherosclerosis." *Annals of the American Thoracic Society* 16 (3): 363–370. doi:10.1513/AnnalsATS.201804-248OC.

Bosco, Liana, Tiziana Notari, Giovanni Ruvolo, Maria C. Roccheri, Chiara Martino, Rosanna Chiappetta, Domenico Carone, et al. 2018. "Sperm DNA fragmentation: An early and reliable marker of air pollution." *Environmental Toxicology and Pharmacology* 58 (March): 243–249. doi:10.1016/j.etap.2018.02.001.

Bowe, Benjamin, Yan Xie, Tingting Li, Yan Yan, Hong Xian, and Ziyad Al-Aly. 2018. "Particulate matter air pollution and the risk of incident CKD and progression to ESRD." *Journal of the American Society of Nephrology* 29 (1): 218–230. doi:10.1681/ASN.2017030253.

Bradshaw, Richard L., Susan Matney, Oren E. Livne, Bruce E. Bray, Joyce A. Mitchell, and Scott P. Narus. 2009. "Architecture of a federated query engine for heterogeneous resources." *AMIA Annual Symposium Proceedings* 2009: 70–74.

Burnett, Nicole, Ram Gouripeddi, Mollie Cummins, Julio Facelli, and Katherine Sward. 2018a. "Towards a molecular basis of exposomic research." In *AMIA 2018 Informatics Summit*, San Francisco, CA. https://knowledge.amia.org/amia-66728-tbi2018-1.4078408/t004-1.4078844.

Burnett, Nicole, Ram Gouripeddi, Naresh Sundar Rajan, Randy Madsen, Peter Mo, and Julio C. Facelli. 2015. "A framework for validating modeled air quality data for use in biomedical research." http://www.airquality.utah.edu/files/2015/01/Burnette_Poster.pdf.

Burnett, Nicole, Ram Gouripeddi, Jingran Wen, Peter Mo, Randy Madsen, Ryan Butcher, Katherine Sward, and Julio C. Facelli. 2017. "Harmonization of sensor metadata and measurements to support exposomic research." In *2016 International Society of Exposure Science*, Research Triangle Park, NC. http://www.intlexposurescience.org/ISES2017.

Burnett, Nicole, Ramkiran Gouripeddi, Julio Facelli, Peter Mo, Randy Madsen, Ryan Butcher, and Katherine Sward. 2018b. "Development of sensor metadata library for exposomic studies." *ISEE Conference Abstracts* 2018 (1). https://ehp.niehs.nih.gov/doi/10.1289/isesisee.2018.P01.0320.

Centers for Disease Control and Prevention. 2019. "Frequently Asked Questions | Social Determinants of Health | NCHHSTP | CDC." https://www.cdc.gov/nchhstp/socialdeterminants/faq.html.

Chatterjee, Madhubanti, Xudong Ge, Yordan Kostov, Leah Tolosa, and Govind Rao. 2014. "A novel approach toward noninvasive monitoring of transcutaneous CO_2." *Medical Engineering and Physics* 36 (1): 136–139. doi:10.1016/j.medengphy.2013.07.001.

Choi, Edwin, and Juhan Sonin. 2019. "Determinants of Health Visualized." https://www.goinvo.com/vision/determinants-of-health/.

Collingwood, Scott, Ramkiran Gouripeddi, Bob Wong, and Katherine Sward. 2018. "Environmental influences on health outcomes-integrating real-time exposure measures from homes into longitudinal cohort studies: Lessons from the field." *ISEE Conference Abstracts* 2018 (1). https://ehp.niehs.nih.gov/doi/10.1289/isesisee.2018.P02.0860.

Collingwood, Scott, Jesse Zmoos, Leon Pahler, Bob Wong, Darrah Sleeth, and Rodney Handy. 2019. “Investigating measurement variation of modified low-cost particle sensors.” *Journal of Aerosol Science* 135 (September): 21–32. doi:10.1016/j.jaerosci.2019.04.017.

Combi, Carlo, Elpida Keravnou-Papailiou, and Yuval Shahar. 2010. *Temporal Information Systems in Medicine*, 1st edition. New York: Springer.

Couchbase. 2019. https://www.couchbase.com/.

Cox, Jennie, Seung-Hyun Cho, Patrick Ryan, Kelechi Isiugo, James Ross, Steven Chillrud, Zheng Zhu, Roman Jandarov, Sergey A. Grinshpun, and Tiina Reponen. 2019. “Combining sensor-based measurement and modeling of PM2.5 and black carbon in assessing exposure to indoor aerosols.” *Aerosol Science and Technology*: 1–13. doi:10.1080/02786826.2019.1608353.

Cummins, Mollie, Ramkiran Gouripeddi, Flory Nkoy, Julio Facelli, and Katherine Sward. 2019a. “Utah PRISMS Center: Sensor-based, data intensive science.” In *Council for the Advancement of Nursing Science 2019 Advanced Methods Conference: The Expanding Science of Sensor Technology in Research*, Washington, DC. https://www.nursingscience.org/events/2019advancedmethods.

Cummins, Mollie, Rima Habre, Ramkiran Gouripeddi, Flory Nkoy, Julio Facelli, and Katherine Sward. 2019b. “Compatibility of sensor study data elements with ECHO.” In *Council for the Advancement of Nursing Science 2019 Advanced Methods Conference: The Expanding Science of Sensor Technology in Research*, Washington, DC. https://www.nursingscience.org/events/2019advancedmethods.

Deelman, Ewa, Karan Vahi, Gideon Juve, Mats Rynge, Scott Callaghan, Philip J. Maechling, Rajiv Mayani, et al. 2015. “Pegasus, a workflow management system for science automation.” *Future Generation Computer Systems* 46 (May): 17–35. doi:10.1016/j.future.2014.10.008.

Deering-Rice, Cassandra E., Darien Shapiro, Erin G. Romero, Chris Stockmann, Tatjana S. Bevans, Quang M. Phan, Bryan L. Stone, et al. 2015. “Activation of transient receptor potential Ankyrin-1 by insoluble particulate material and association with asthma.” *American Journal of Respiratory Cell and Molecular Biology* 53 (6): 893–901. doi:10.1165/rcmb.2015-0086OC.

Deering-Rice, Cassandra E., Chris Stockmann, Erin G. Romero, Zhenyu Lu, Darien Shapiro, Bryan L. Stone, Bernhard Fassl, et al. 2016. “Characterization of Transient Receptor Potential Vanilloid-1 (TRPV1) Variant Activation by Coal Fly Ash Particles and Associations with Altered Transient Receptor Potential Ankyrin-1 (TRPA1) Expression and Asthma.” *Journal of Biological Chemistry* 291 (48): 24866–24879. doi:10.1074/jbc.M116.746156.

Dittrich, Yvonne. 2014. “Software engineering beyond the project: Sustaining software ecosystems.” *Information and Software Technology, Special Issue on Software Ecosystems* 56 (11): 1436–1456. doi:10.1016/j.infsof.2014.02.012.

Erickson, Larry E., and Merrisa Jennings. 2017. “Energy, transportation, air quality, climate change, health nexus: Sustainable energy is good for our health.” *AIMS Public Health* 4 (1): 47–61. doi:10.3934/publichealth.2017.1.47.

Farahani, Shahin. 2011. *ZigBee Wireless Networks and Transceivers*. Oxford: Newnes.

Friedman, Charles, Joshua Rubin, Jeffrey Brown, Melinda Buntin, Milton Corn, Lynn Etheredge, Carl Gunter, et al. 2014. “Toward a science of learning systems: A research agenda for the high-functioning learning health system.” *Journal of the American Medical Informatics Association*. doi:10.1136/amiajnl-2014-002977.

Gładka, Anna, Joanna Rymaszewska, and Tomasz Zatoński. 2018. “Impact of air pollution on depression and suicide.” *International Journal of Occupational Medicine and Environmental Health* 31 (6): 711–721. doi:10.13075/ijomeh.1896.01277.

Gouripeddi, Ramkiran. 2016. “An informatics architecture for an exposome.” In *American Medical Informatics Association Spring 2016*. https://www.amia.org/jointsummits2016/implementation-informatics-track-sessions.

Gouripeddi, Ramkiran. 2019. OpenFurther, http://Openfurther.Org/ and Utah Center for Clinical and Translational Science, http://openfurther.org/.

Gouripeddi, Ramkiran, Nicole Burnett, Mollie Cummins, Julio Facelli, and Katherine Sward. 2017. “A conceptual representation of exposome in translational research.” In *AMIA 2017 Annual Symposium*, Washington, DC. https://amia2017.zerista.com/speaker/print?order=company&speaker_page=16.

Gouripeddi, Ramkiran, Nicole Burnett, and Julio Facelli. 2015. “Data, modeling, uncertainty and integration: The informatics of an air quality exposome.” In *2015 International Society of Exposure Science Annual Meeting*, Henderson, NV. http://www.ises2015.org/Images/ISES-progam-DRAFT1.pdf.

Gouripeddi, Ramkiran, Scott C. Collingwood, Bob Wong, Mollie Cummins, Julio Facelli, and Katherine Sward. 2018. “The Utah PRISMS informatics ecosystem: An infrastructure for generating and utilizing exposomes for translational research.” In *Total Exposure Health Conference 2018*, Bethesda, MD. http://www.cvent.com/events/total-exposure-health-2018/event-summary-f5ae5c586a414ceb84a7fb594789dd7b.aspx?lang=en.

Gouripeddi, Ramkiran, Rima Habre, and PRISMS Data Modeling Working Group. 2019. “Developing a specification for representing exposure health semantics.” In *The International Societies of Exposure Science (ISES) and Indoor Air Quality and Climate (ISIAQ) 2019 Conference*, Kaunas, Lithuania.

Gouripeddi, Ramkiran, Andrew Miller, Karen Eilbeck, Katherine Sward, and Julio C. Facelli. 2019a. “Systematically integrating microbiomes and exposomes for translational research.” *Journal of Clinical and Translational Science* 3 (s1): 29–30. doi:10.1017/cts.2019.71.

Gouripeddi, Ramkiran, N Dustin Schultz, Richard L. Bradshaw, Peter Mo, Randy Madsen, Phillip B. Warner, Bernard A. LaSalle, and Julio C Facelli. 2013. “FURTHeR: An infrastructure for clinical, translational and comparative effectiveness research.” In *American Medical Informatics Association* 2013 *Annual Symposium*, Washington, DC. http://knowledge.amia.org/amia-55142-a2013e-1.580047/t-10-1.581994/f-010-1.581995/a-184-1.582011/ap-247-1.582014.

Gouripeddi, Ramkiran, Le-Thuy Tran, Albert Lund, Randy Madsen, Julio Facelli, and Sward Katherine. 2019b. “The Utah PRISMS informatics ecosystem: An infrastructure for generating and utilizing exposomes for translational research.” In *AMIA* 2019 *Informatics Summit*, San Francisco, CA. https://informaticssummit2019.zerista.com/event/member/543073, Presentation: https://uofu.box.com/s/t7ehydznuwx5muka6ctof1dcu7aecfcv.

Gouripeddi, Ramkiran, Le-Thuy Tran, Randy Madsen, Tanvi Gangadhar, Peter Mo, Nicole Burnett, Ryan Butcher, Katherine Sward, and Julio Facelli. 2019c. “An architecture for metadata-driven integration of heterogeneous sensor and health data for translational exposomic research.” In *2019 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI) (IEEE BHI* 2019*)*, Chicago, IL.

Gouripeddi, Ramkiran, Phillip B. Warner, Peter Mo, James E. Levin, Rajendu Srivastava, Samir S. Shah, David de Regt, Eric Kirkendall, Jonathan Bickel, and E. Kent Korgenski. 2012. “Federating clinical data from six pediatric hospitals: Process and initial results for microbiology from the PHIS+ consortium.” In *AMIA Annual Symposium Proceedings*, 281. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540481/.

Grafana - The Open Platform for Analytics and Monitoring. 2019. https://grafana.com/.

Groat, Danielle, Ram Gouripeddi, Yu Keui Lin, and Julio C. Facelli. 2019. “Measuring the autonomic nervous system for translational research: Identification of non-invasive methods.” *Journal of Clinical and Translational Science* 3 (s1): 28–28. doi:10.1017/cts.2019.68.

Habre, Rima, Jose-Luis Ambite, Alex Bui, Mollie Cummins, Michael Dellarco, Sandrah Eckel, Ramkiran Gouripeddi, and Katherine Sward. 2016. "Pediatric Research Using Integrated Sensor Monitoring Systems, Data Modeling Working Group." In *26th Annual International Society of Exposure Science (ISES) Meeting*, Utrecht, The Netherlands. https://www.intlexposurescience.org/Public/Annual_Meeting/Past_Meetings/26th_Annual_Meeting/Public/Annual_Meeting/PastMeetings/2016_Meeting.aspx?hkey=9ae9d30a-6e0b-4177-98d5-95e58a1f64fb.

Harris, Paul A., Robert Taylor, Robert Thielke, Jonathon Payne, Nathaniel Gonzalez, and Jose G. Conde. 2009. "Research Electronic Data Capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support." *Journal of Biomedical Informatics* 42 (2): 377–381. doi:10.1016/j.jbi.2008.08.010.

Hichwa, Paul, and Cristina E Davis. 2018. "Modern application potential of miniature chemical sensors." In Kevin Yallup and Laura Basiricò (Eds.), *Sensors for Diagnostics and Monitoring*, 295–314. Boca Raton, FL: CRC Press.

Home Assistant. 2019. https://www.home-assistant.io/.

Horel, J., M. Splitt, L. Dunn, J. Pechmann, B. White, C. Ciliberti, S. Lazarus, J. Slemmer, D. Zaff, and J. Burks. 2002. "Mesowest: Cooperative mesonets in the Western United States." *Bulletin of the American Meteorological Society* 83 (2): 211–226. doi:10.1175/1520-0477(2002)083<0211:MCMITW>2.3.CO;2.

Hunkeler, U., H. L. Truong, and A. Stanford-Clark. 2008. "MQTT-S: A publish/subscribe protocol for wireless sensor networks." In *2008 3rd International Conference on Communication Systems Software and Middleware and Workshops (COMSWARE'08)*, 791–798. doi:10.1109/COMSWA.2008.4554519.

InfluxData (InfluxDB) | Time Series Database Monitoring & Analytics. 2019. https://www.influxdata.com/.

Jacobs, David E., Jonathan Wilson, Sherry L. Dixon, Janet Smith, and Anne Evens. 2009. "The Relationship of housing and population health: A 30-year retrospective analysis." *Environmental Health Perspectives* 117 (4): 597–604. doi:10.1289/ehp.0800086.

Jansen, Slinger. 2014. "Measuring the health of open source software ecosystems: Beyond the scope of project health." *Information and Software Technology, Special Issue on Software Ecosystems*, 56 (11): 1508–1519. doi:10.1016/j.infsof.2014.04.006.

Jansen, Slinger, Sjaak Brinkkemper, and Michael A. Cusumano. 2013. *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*. Cheltenham, UK and Northampton, MA: Edward Elgar Publisher.

Jansen, Slinger, Anthony Finkelstein, and Sjaak Brinkkemper. 2009. "A sense of community: A research agenda for software ecosystems." In *2009 31st International Conference on Software Engineering: Companion Volume*, 187–190, Vancouver, BC, Canada: IEEE. doi:10.1109/ICSE-COMPANION.2009.5070978.

Kasera, Sneha Kumar. 2019. "Powder: Platform for open wireless data-driven experimental research." https://powderwireless.net/.

Kaskel, Frederick, Daniel Batlle, Srinivasan Beddhu, John Daugirdas, Harold Feldman, Maria Ferris, Lawrence Fine, et al. 2014. "Improving CKD therapies and care: A national dialogue." *Clinical Journal of the American Society of Nephrology* 9 (4): 815–817. doi:10.2215/CJN.12631213.

Kelly, K.E., J. Whitaker, A. Petty, C. Widmer, A. Dybwad, D. Sleeth, R. Martin, and A. Butterfield. 2017. "Ambient and laboratory evaluation of a low-cost particulate matter sensor." *Environmental Pollution* 221 (February): 491–500. doi:10.1016/j.envpol.2016.12.039.

Kiossoglou, Phillip, Ann Borda, Kathleen Gray, Ferdinando Martin-Sanchez, Karin Verspoor, and Guillermo H. Lopez Campos. 2017. "How well do existing ontologies represent exposome literature?" In *AMIA 2017 Informatics Summit*, San Francisco, CA. https://knowledge.amia.org/amia-64484-tbi2017-1.3520073/t004-1.3520463.

Kitt, Jay, Randall Martin, Krishna Sundar, Cheryl Pirozzi, Ramkiran Gouripeddi, Joel Harris, and Facelli Julio. 2019. "Informatics and chemistry: Examining the impact of particulate matter chemical composition on pulmonary disease outcomes." In *The Air We Breathe: A Multidisciplinary Perspective on Air Quality*, Salt Lake City, UT.

Kuc, Rafal, and Marek Rogozinski. 2013. *Elasticsearch Server*. Birmingham: Packt Publishing Ltd.

Kukkapalli, R., N. Banerjee, R. Robucci, and Y. Kostov. 2016. "Micro-radar wearable respiration monitor." In *2016 IEEE Sensors*, 1–3. doi:10.1109/ICSENS.2016.7808741.

Landrigan, Philip J., Richard Fuller, Nereus J. R. Acosta, Olusoji Adeyi, Robert Arnold, Niladri (Nil) Basu, Abdoulaye Bibi Baldé, et al. 2018. "The Lancet Commission on pollution and health." *The Lancet* 391 (10119): 462–512. doi:10.1016/S0140-6736(17)32345-0.

Lee, Byeong-Jae, Bumseok Kim, and Kyuhong Lee. 2014. "Air pollution exposure and cardiovascular disease." *Toxicological Research* 30 (2): 71–75. doi:10.5487/TR.2014.30.2.071.

Lee, H., and K. Ke. 2018. "Monitoring of large-area IoT sensors using a LoRa wireless mesh network system: Design and evaluation." *IEEE Transactions on Instrumentation and Measurement* 67 (9): 2177–2187. doi:10.1109/TIM.2018.2814082.

Leiser, Claire L., Heidi A. Hanson, Kara Sawyer, Jacob Steenblik, Ragheed Al-Dulaimi, Troy Madsen, Karen Gibbins, et al. 2019. "Acute effects of air pollutants on spontaneous pregnancy loss: A case-crossover study." *Fertility and Sterility* 111 (2): 341–347. doi:10.1016/j.fertnstert.2018.10.028.

Li, Baichen, Quan Dong, R. Scott Downen, Nam Tran, J. Hunter Jackson, Dinesh Pillai, Mona Zaghloul, and Zhenyu Li. 2019. "A wearable IoT aldehyde sensor for pediatric asthma research and management." *Sensors and Actuators B: Chemical* 287: 584–594.

Lioy, Paul, and Clifford Weisel. 2014. *Exposure Science: Basic Principles and Applications*, 1st edition. Amsterdam and Boston, MA: Academic Press.

Livne, Oren E., N. Dustin Schultz, and Scott P. Narus. 2011. "Federated querying architecture with clinical & translational health IT application." *Journal of Medical Systems* 35 (5): 1211–1224. doi:10.1007/s10916-011-9720-3.

Lopez-Campos, Guillermo, Christopher Hawthorn, Philip Kiossoglou, Ramkiran Gouripeddi, Ann Borda, Kathleen Gray, Fernando Martin-Sanchez, and Karin Verspoor. 2019. "Characterising the scope of exposome research: A generalisable approach." In *2019 Exposome Symposium*, Brescia, Italy.

Lopez-Campos, G., M. Merolli, and F. Martin-Sanchez. 2017. "Biomedical informatics and the digital component of the exposome." *Studies in Health Technology and Informatics* 245: 496–500.

Lund, Albert, Nicole Burnett, Ram Gouripeddi, and Julio Facelli. 2017. "An agent-based model for estimating human activity patterns on the wasatch front." In *Air Quality: Science for Solutions*, Salt Lake City, UT. http://harbor.weber.edu/Airqualityscience/abstracts/default.html.

Lund, Albert, Ram Gouripeddi, Nicole Burnett, Le-Thuy Tran, Peter Mo, Randy Madsen, Mollie Cummins, Katherine Sward, and Julio Facelli. 2018. "Enabling reproducible computational modeling: The Utah PRISMS Ecosystem." In *Conference: Building Research Integrity Through Reproducibility*, Salt Lake City, UT. campusguides.lib.utah.edu/UtahRR18/abstracts.

Lund, Albert M., Ramkiran Gouripeddi, and Julio C. Facelli. 2019a. "Generation and classification of activity sequences for spatiotemporal modeling of human populations." ArXiv:1911.05476 [Cs], November. http://arxiv.org/abs/1911.05476.

Lund, Albert, Ramkiran Gouripeddi, and Julio C. Facelli. 2019b. "STHAM: An agent based model for simulating human exposure across high resolution spatiotemporal domains." *Journal of Exposure Science and Environmental Epidemiology*, In press. doi.org/10.1038/s41370-020-0216-4.

Lundrigan, Philip. 2019. VDL-PRISM. https://github.com/VDL-PRISM.

Lundrigan, Philip, Sneha Kumar Kasera, and Neal Patwari. 2018. "STRAP: Secure TRansfer of Association Protocol." In *2018 27th International Conference on Computer Communication and Networks (ICCCN)*, 1–9. doi:10.1109/ICCCN.2018.8487333.

Lundrigan, Philip, Kyeong Min, Neal Patwari, Sneha Kasera, Kerry Kelly, Jimmy Moore, Miriah Meyer, et al. 2017. "EpiFi: An In-Home Sensor Network Architecture for Epidemiological Studies." *ArXiv:1709.02233 [Cs]*, September. http://arxiv.org/abs/1709.02233.

Lundrigan, Philip, Kyeong T. Min, Neal Patwari, Sneha Kumar Kasera, Kerry Kelly, Jimmy Moore, Miriah Meyer, et al. 2018. "EpiFi: An in-Home IoT Architecture for Epidemiological Deployments." In *2018 IEEE 43rd Conference on Local Computer Networks Workshops (LCN Workshops)*, 30–37. doi:10.1109/LCNW.2018.8628482.

Lungu, Mircea. 2009. "Reverse Engineering Software Ecosystems," September, 208.

Luxton, David D., Robert A. Kayl, and Matthew C. Mishkind. 2012. "MHealth data security: The need for HIPAA-compliant standardization." *Telemedicine and E-Health* 18 (4): 284–288. doi:10.1089/tmj.2011.0180.

Manikas, Konstantinos, and Klaus Marius Hansen. 2013. "Software ecosystems: A systematic literature review." *Journal of Systems and Software* 86 (5): 1294–1306. doi:10.1016/j.jss.2012.12.026.

Martin-Sanchez, Fernando, Kathleen Gray, Riccardo Bellazzi, and Guillermo Lopez-Campos. 2014. "Exposome informatics: Considerations for the design of future biomedical research information systems." *Journal of the American Medical Informatics Association* 21 (3): 386–390. doi:10.1136/amiajnl-2013-001772.

Mattingly, Carolyn J., Rebecca Boyles, Cindy P. Lawler, Astrid C. Haugen, Allen Dearry, and Melissa Haendel. 2016. "Laying a Community-Based Foundation for Data-Driven Semantic Standards in Environmental Health Sciences." *Environmental Health Perspectives* 124 (8). doi:10.1289/ehp.1510438.

McCartney, Mitchell M., Yuriy Zrodnikov, Alexander G. Fung, Michael K. LeVasseur, Josephine M. Pedersen, Konstantin O. Zamuruyev, Alexander A. Aksenov, Nicholas J. Kenyon, and Cristina E. Davis. 2017. "An easy to manufacture micro gas preconcentrator for chemical sensing applications." *ACS Sensors* 2 (8): 1167–1174. doi:10.1021/acssensors.7b00289.

McGinnis, J. Michael, Pamela Williams-Russo, and James R. Knickman. 2002. "The case for more active policy attention to health promotion." *Health Affairs (Project Hope)* 21 (2): 78–93. doi:10.1377/hlthaff.21.2.78.

McMillan, Nancy J., David M. Holland, Michele Morara, and Jingyu Feng. 2010. "Combining numerical model output and particulate data using Bayesian space–time modeling." *Environmetrics* 21 (1): 48–65. doi:10.1002/env.984.

McMullen, C. K., J. S. Ash, D. F. Sittig, A. Bunce, K. Guappone, R. Dykstra, J. Carpenter, J. Richardson, and A. Wright. 2011. "Rapid assessment of clinical information systems in the healthcare setting: An efficient method for time-pressed evaluation." *Methods of Information in Medicine* 50 (4): 299–307. doi:10.3414/ME10-01-0042.

Mehta, Maitrey, Danielle Groat, Yu Kuei Lin, Ramkiran Gouripeddi, and Julio Facelli. 2019. "Classifying impaired awareness of hypoglycemia with convolutional neural networks." In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI) (IEEE BHI 2019)*, Chicago, IL.

Messerschmitt, David G., and Clemens Szyperski. 2003. *Software Ecosystem: Understanding an Indispensable Technology and Industry.* Cambridge, MA: The MIT Press.

Miller, Gary W. 2013. *The Exposome: A Primer*, 1st edition. Amsterdam and Boston, MA: Academic Press.

Miller, Gary W., and Dean P. Jones. 2014. "The nature of nurture: Refining the definition of the exposome." *Toxicological Sciences* 137 (1): 1–2. doi:10.1093/toxsci/kft251.

Min, Kyeong T., Philip Lundrigan, Katherine Sward, Scott C. Collingwood, and Neal Patwari. 2018. “Smart home air filtering system: A randomized controlled trial for performance evaluation.” *Smart Health*. doi:10.1016/j.smhl.2018.07.009.

Miri, Mohammad, Milad Nazarzadeh, Ahmad Alahabadi, Mohammad Hassan Ehrampoush, Abolfazl Rad, Mohammad Hassan Lotfi, Mohammad Hassan Sheikhha, Mohammad Javad Zare Sakhvidi, Tim S. Nawrot, and Payam Dadvand. 2019. “Air pollution and telomere length in adults: A systematic review and meta-analysis of observational studies.” *Environmental Pollution (Barking, Essex: 1987)* 244 (January): 636–647. doi:10.1016/j.envpol.2018.09.130.

Mirowsky, Jaime, Christina Hickey, Lori Horton, Martin Blaustein, Karen Galdanes, Richard E. Peltier, Steven Chillrud, et al. 2013. “The effect of particle size, location and season on the toxicity of urban and rural particulate matter.” *Inhalation Toxicology* 25 (13): 747–757. doi:10.3109/08958378.2013.846443.

Mitchell, L., E. Crosman, B. Fasoli, L. Leclair-Marzolf, A. Jacques, J. Horel, J. C. Lin, D. R. Bowling, and J. R. Ehleringer. 2015. “Spatiotemporal patterns of urban trace gases and pollutants observed with a light rail vehicle platform in Salt Lake City, UT.” *AGU Fall Meeting Abstracts* 23 (December): A23P-06.

Mitchell, Logan E., Erik T. Crosman, Alexander A. Jacques, Benjamin Fasoli, Luke Leclair-Marzolf, John Horel, David R. Bowling, James R. Ehleringer, and John C. Lin. 2018. “Monitoring of greenhouse gases and pollutants across an urban area using a light-rail public transit platform.” *Atmospheric Environment* 187 (August): 9–23. doi:10.1016/j.atmosenv.2018.05.044.

Mo, Peter, N. Dustin Schultz, Richard L. Bradshaw, Ryan Butcher, R. Gouripeddi, Phillip B. Warner, Randy Madsen, Bernard A. LaSalle, and Julio C. Facelli. 2014. “Real-time federated data translations using metadata-driven XQuery.” In *AMIA 2014 Summit on Clinical Research Informatics*, San Francisco, CA. http://knowledge.amia.org/amia-56636-cri2014-1.977698/t-004-1.978136/a-089-1.978209/a-089-1.978210/ap-085-1.978211.

MongoDB | The Most Popular Database for Modern Apps. 2019. https://www.mongodb.com/index.

Moore, Jimmy, Pascal Goffin, Miriah Meyer, Philip Lundrigan, Neal Patwari, Katherine Sward, and Jason Wiese. 2018. “Managing in-home environments through sensing, annotating, and visualizing air quality data.” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2 (3): 28. doi:10.1145/3264938.

Morgan, Taylor, Ramkiran Gouripeddi, and Katherine Sward. 2019. “Characterizing metadata of e-devices for interventional translational research.” In *AMIA 2019 Informatics Summit*, San Francisco, CA. https://www.amia.org/summit2019.

Narkhede, Neha, Gwen Shapira, and Todd Palino. 2017. *Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale*, 1st edition. Sebastopol, CA: O’Reilly Media.

National Institute of Environmental Health Sciences (NIEHS). 2012. “National Institute of Environmental Health Sciences Strategic Plan 2012–2017.” https://www.niehs.nih.gov/about/strategicplan/strategicplan2012/index.cfm.

National Institute of Environmental Health Sciences (NIEHS). 2018. “National Institute of Environmental Health Sciences Strategic Plan 2018–2023: Advancing Environmental Health Sciences, Improving Health.” https://www.niehs.nih.gov/about/strategicplan/index.cfm.

National Institute for Occupational Safety and Health (NIOSH). 2018. CDC - Exposome and Exposomics - NIOSH Workplace Safety and Health Topic. https://www.cdc.gov/niosh/topics/exposome/default.html.

Neo4j Graph Platform – The Leader in Graph Databases. 2019. https://neo4j.com/.

Newton, John N., Adam D. M. Briggs, Christopher J. L. Murray, Daniel Dicker, Kyle J. Foreman, Haidong Wang, Mohsen Naghavi, et al. 2015. "Changes in health in England, with analysis by English regions and areas of deprivation, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013." *Lancet (London, England)* 386 (10010): 2257–2274. doi:10.1016/S0140-6736(15)00195-6.

Nkoy, Flory L., Bryan L. Stone, Bernhard A. Fassl, Karmella Koopmeiners, Sarah Halbern, Eun H. Kim, Justin Poll, Joseph W. Hales, Dillon Lee, and Christopher G. Maloney. 2012. "Development of a novel tool for engaging children and parents in asthma self-management." *AMIA Annual Symposium Proceedings* 2012 (November): 663–672.

Nkoy, Flory L., Bryan L. Stone, Bernhard A. Fassl, Derek A. Uchida, Karmella Koopmeiners, Sarah Halbern, Eun H. Kim, et al. 2013. "Longitudinal validation of a tool for asthma self-monitoring." *Pediatrics* 132 (6): e1554–e1561. doi:10.1542/peds.2013–1389.

Object Management Group (OMG). 2019. https://www.service-architecture.com/articles/web-services/object_management_group_omg.html.

OrientDB - Distributed Multi-Model and Graph Database. 2019. http://orientdb.com/orientdb/.

Pirozzi, Cheryl S., Barbara E. Jones, James A. VanDerslice, Yue Zhang, Robert Paine, and Nathan C. Dean. 2017. "Short-term air pollution and incident pneumonia. A case–crossover study." *Annals of the American Thoracic Society* 15 (4): 449–459. doi:10.1513/AnnalsATS.201706-495OC.

Pirozzi, Cheryl, Anne Sturrock, Hsin-Yi Weng, Tom Greene, Mary Beth Scholand, Richard Kanner, and Robert Paine Iii. 2015. "Effect of naturally occurring ozone air pollution episodes on pulmonary oxidative stress and inflammation." *International Journal of Environmental Research and Public Health* 12 (5): 5061–5075. doi:10.3390/ijerph120505061.

Pollak, Bill, ed. 2006. *Ultra-Large-Scale Systems the Software Challenge of the Future*. Pittsburgh, PA: Software Engineering Institute Carnegie Mellon.

Pollock, Jenna, Lu Shi, and Ronald W. Gimbel. 2017. "Outdoor environment and pediatric asthma: An update on the evidence from North America." *Canadian Respiratory Journal* 2017: 8921917. doi:10.1155/2017/8921917.

Popp, Karl, and Ralf Meyer. 2010. *Profit from Software Ecosystems*. Norderstedt: Books on Demand.

Prüss-Üstün, Annette, J. Wolf, C. Corvalán, Robert Bos, and M. Neira. 2016. *Preventing Disease through Healthy Environments: A Global Assessment of the Burden of Disease from Environmental Risks*, 2nd edition. Geneva, Switzerland: World Health Organization.

PurpleAir.Org. 2019. http://www.purpleair.org/.

Riches, Naomi, Ramkiran Gouripeddi, Kevin Perry, and Rod Handy. 2019. "Informatics methods for characterizing multi-pollutant air quality exposure on health." In *The Air we Breathe: A Multidisciplinary Perspective on Air Quality*, Salt Lake City, UT.

Richtel, Matt. 2019. *An Elegant Defense: The Extraordinary New Science of the Immune System: A Tale in Four Lives*, Larger Print edition. New York: HarperLuxe.

Robinson, Ian, Jim Webber, and Emil Eifrem. 2015. *Graph Databases: New Opportunities for Connected Data*, 2nd edition. Beijing: O'Reilly Media.

Röhrig, Bernd, Jean-Baptist du Prel, Daniel Wachtlin, and Maria Blettner. 2009. "Types of study in medical research." *Deutsches Arzteblatt International* 106 (15): 262–268. doi:10.3238/arztebl.2009.0262.

Santibáñez-Andrade, Miguel, Ericka Marel Quezada-Maldonado, Álvaro Osornio-Vargas, Yesennia Sánchez-Pérez, and Claudia M. García-Cuellar. 2017. "Air pollution and genomic instability: The role of particulate matter in lung carcinogenesis." *Environmental Pollution* 229 (October): 412–422. doi:10.1016/j.envpol.2017.06.019.

Sarbhai, Aarushi, Ramkiran Gouripeddi, Philip Lundrigan, Pavithra Chidambaram, Aakansha Saha, Julio C. Facelli, Katherine Sward, and Sneha Kumar Kasera. 2019. "Managing sensor metadata in exposomic studies using blockchain." In *The Air We Breathe: A Multidisciplinary Perspective on Air Quality*, Salt Lake City, UT.

Schivo, Michael, Felicia Seichter, Alexander A. Aksenov, Alberto Pasamontes, Daniel J. Peirano, Boris Mizaikoff, Nicholas J. Kenyon, and Cristina E. Davis. 2013. "A mobile instrumentation platform to distinguish airway disorders." *Journal of Breath Research* 7 (1): 017113. doi:10.1088/1752-7155/7/1/017113.

Shiffman, Saul, Arthur A. Stone, and Michael R. Hufford. 2008. "Ecological momentary assessment." *Annual Review of Clinical Psychology* 4 (1): 1–32. doi:10.1146/annurev.clinpsy.3.022806.091415.

Singleton, John Robinson, Billie Bixby, James W. Russell, Eva L. Feldman, Amanda Peltier, Jonathan Goldstein, James Howard, and A. Gordon Smith. 2008. "The Utah early neuropathy scale: A sensitive clinical scale for early sensory predominant neuropathy." *Journal of the Peripheral Nervous System* 13 (3): 218–227. doi:10.1111/j.1529-8027.2008.00180.x.

Singleton, John R., Robin L. Marcus, Justin E. Jackson, Margaret K. Lessard, Timothy E. Graham, and Albert G. Smith. 2014. "Exercise increases cutaneous nerve density in diabetic patients without neuropathy." *Annals of Clinical and Translational Neurology* 1 (10): 844–849. doi:10.1002/acn3.125.

Stanford, Joseph, and Scott Collingwood. 2017. "Utah's Children's Project." https://healthcare.utah.edu/.

Stripelis, D., J. L. Ambite, Y. Chiang, S. P. Eckel, and R. Habre. 2017. "A scalable data integration and analysis architecture for sensor data of pediatric asthma." In *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, 1407–1408. doi:10.1109/ICDE.2017.198.

Sundar, Krishna M., Sarah E. Daly, and Alika M. Willis. 2013. "A longitudinal study of CPAP therapy for patients with chronic cough and obstructive sleep apnoea." *Cough* 9 (1): 19. doi:10.1186/1745-9974-9-19.

Sward, Katherine. 2019. "Making the connection: Integrated sensor systems research." In *52nd Annual Communicating Nursing Research Conference*, San Diego, CA. https://www.winursing.org/2019-researchconference/.

Sward, Katherine, Alex Bui, Jose-Luis Ambite, Michael Dellarco, Julio Facelli, and Frank D. Gilliland. 2016. "Pediatric Research Using Integrated Sensor Monitoring Systems (PRISMS): Applying sensor technology and informatics to better understand asthma." In *2016 American Medical Informatics Association Annual Symposium*, Chicago, IL. https://www.amia.org/amia2016.

Sward, Katherine, and Julio Facelli. 2016. "Pediatric Research Using Integrated Sensor Monitoring Systems (PRISMS): Utah Informatics Center Federated Integration Architecture." In *2016 International Society of Exposure Science Annual Meeting*, Utrecht, The Netherlands. https://ises2016.org/.

Sward, Katherine, Neal Patwari, Ramkiran Gouripeddi, and Julio Facelli. 2017. "An infrastructure for generating exposomes: Initial lessons from the Utah PRISMS platform." In *2017 International Society of Exposure Science Annual Meeting,* Research Triangle Park, NC. https://intlexposurescience.org/ISES2017/default.aspx.

Tarlov, A. R. 1999. "Public policy frameworks for improving population health." *Annals of the New York Academy of Sciences* 896: 281–293.

Tiase, Victoria, Ram Gouripeddi, Nicole Burnett, Ryan Butcher, Peter Mo, Mollie Cummins, and Katherine Sward. 2018. "Advancing study metadata models to support an exposomic informatics infrastructure." In *2018 ISES Annual Meeting*, *Ottawa*, Canada. http://www.eiseverywhere.com/ehome/294696/638649/?&t=8c531cecd4bb0a5efc6a0045f5bec0c3.

Twilio: Communication APIs for SMS, Voice, Video and Authentication. 2019. Accessed June 1. https://www.twilio.com.

United Nations International Children's Emergency Fund. 2016. "Clear the Air for Children." https://www.unicef.org/publications/index_92957.html.

US Bureau of Labor Statistics. 2017. *American Time Use Survey*. https://www.bls.gov/tus/.

US Census Bureau. 2010. *Decennial Census by Decades*. https://www.census.gov/programs-surveys/decennial-census/decade.html.

US Census Bureau. 2017. *American Community Survey (ACS)*. https://www.census.gov/programs-surveys/acs.

US Census Bureau Center for Economic Studies. 2016. US Census Bureau Center for Economic Studies Publications and Reports Page. https://lehd.ces.census.gov/data/.

US Department of Commerce, NOAA. 2019. *National Weather Service*. http://www.weather.gov/.

US Environmental Protection Agency. 2005. "Revision to the guideline on air quality models: Adoption of a Preferred general purpose (flat and complex terrain) dispersion model and other revisions." Federal Register. https://www.federalregister.gov/documents/2005/11/09/05-21627/revision-to-the-guideline-on-air-quality-models-adoption-of-a-preferred-general-purpose-flat-and.

US Environmental Protection Agency. 2016a. "Model Clearinghouse Information Storage and Retrieval System | Support Center for Regulatory Atmospheric Modeling | Technology Transfer Network | US EPA." https://cfpub.epa.gov/oarweb/MCHISRS/.

US Environmental Protection Agency. 2016b. "Air Quality Data for the CDC National EPHT Network | Human Exposure and Atmospheric Sciences | US EPA." https://archive.epa.gov/heasd/archive-heasd/web/html/cdc.html.

US Environmental Protection Agency. 2016c. "Support Center for Regulatory Atmospheric Modeling (SCRAM)." *Data and Tools*. https://www.epa.gov/scram.

Vercellino, Robert J., Darrah K. Sleeth, Rodney G. Handy, Kyeong T. Min, and Scott C. Collingwood. 2018. "Laboratory evaluation of a low-cost, real-time, aerosol multi-sensor." *Journal of Occupational and Environmental Hygiene* 15 (7): 559–567. doi:10.1080/15459624.2018.1468565.

Wang, Di, and Nongjian Tao. 2017. A New Take on Asthma Sensors Offers Breath of Fresh Air: The State Press. https://www.statepress.com/article/2017/04/spscience-prisms-air-quality-watch.

Weinreich, Gerhard, Thomas E. Wessendorf, Noreen Pundt, Gudrun Weinmayr, Frauke Hennig, Susanne Moebus, Stefan Möhlenkamp, et al. 2015. "Association of short-term ozone and temperature with sleep disordered breathing." *The European Respiratory Journal* 46 (5): 1361–1369. doi:10.1183/13993003.02255-2014.

Wild, Christopher Paul. 2005. "Complementing the genome with an 'exposome': The outstanding challenge of environmental exposure measurement in molecular epidemiology." *Cancer Epidemiology and Prevention Biomarkers* 14 (8): 1847–1850. doi:10.1158/1055-9965.EPI-05-0456.

Wild, Christopher Paul. 2012. "The exposome: From concept to utility." *International Journal of Epidemiology* 41 (1): 24–32. doi:10.1093/ije/dyr236.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR guiding principles for scientific data management and stewardship." *Scientific Data* 3 (March): 160018. doi:10.1038/sdata.2016.18.

Williams, R., R. Duvall, V. Kilaru, G. Hagler, L. Hassinger, K. Benedict, J. Rice, et al. 2019. "Deliberating performance targets workshop: Potential paths for emerging PM2.5 and O_3 air sensor progress." *Atmospheric Environment* X 2 (April): 100031. doi:10.1016/j.aeaoa.2019.100031.

Wnuk, Krzysztof, Per Runeson, Matilda Lantz, and Oskar Weijden. 2014. "Bridges and barriers to hardware-dependent software ecosystem participation: A case study." *Information and Software Technology, Special issue on Software Ecosystems*, 56 (11): 1493–1507. doi:10.1016/j.infsof.2014.05.015.

World Health Organization. 2014. "WHO | 7 Million Premature Deaths Annually Linked to Air Pollution." https://www.who.int/mediacentre/news/releases/2014/air-pollution/en/.

World Health Organization. 2016a. "WHO | By Category | Deaths Attributable to the Environment: Data by Country." http://apps.who.int/gho/data/node.main.162.

World Health Organization. 2016b. "An Estimated 12.6 Million Deaths Each Year Are Attributable to Unhealthy Environments." https://www.who.int/news-room/detail/15-03-2016-an-estimated-12-6-million-deaths-each-year-are-attributable-to-unhealthy-environments.

World Health Organization. 2017. WHO | Inheriting a Sustainable World: Atlas on Children's Health and the Environment. http://www.who.int/ceh/publications/inheriting-a-sustainable-world/en/.

Yang, Yang, Zengliang Ruan, Xiaojie Wang, Yin Yang, Tonya G. Mason, Hualiang Lin, and Linwei Tian. 2019. "Short-term and long-term exposures to fine particulate matter constituents and health: A systematic review and meta-analysis." *Environmental Pollution* 247 (April): 874–882. doi:10.1016/j.envpol.2018.12.060.

Yanosky, Jeff D., Christopher J. Paciorek, Francine Laden, Jaime E. Hart, Robin C. Puett, Duanping Liao, and Helen H. Suh. 2014. "Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors." *Environmental Health* 13 (1): 63. doi:10.1186/1476-069X-13-63.

Yassein, M. B., W. Mardini, and A. Khalil. 2016. "Smart homes automation using Z-wave protocol." In *2016 International Conference on Engineering MIS (ICEMIS)*, 1–6. doi:10.1109/ICEMIS.2016.7745306.

Zanobetti, Antonella, Susan Redline, Joel Schwartz, Dennis Rosen, Sanjay Patel, George T. O'Connor, Michael Lebowitz, Brent A. Coull, and Diane R. Gold. 2010. "Associations of PM10 with sleep and sleep-disordered breathing in adults from seven U.S. Urban Areas." *American Journal of Respiratory and Critical Care Medicine* 182 (6): 819–825. doi:10.1164/rccm.200912-1797OC.

Zeger, S. L., D. Thomas, F. Dominici, J. M. Samet, J. Schwartz, D. Dockery, and A. Cohen. 2000. "Exposure measurement error in time-series studies of air pollution: Concepts and consequences." *Environmental Health Perspectives* 108 (5): 419–426. doi:10.1289/ehp.00108419.